

СТЮАРТ РИЧИ

STUART RITCHIE

SCIENCE FICTIONS

EXPOSING FRAUD,
BIAS, NEGLIGENCE AND
HYPE IN SCIENCE

СТЮАРТ РИЧИ

НАУКО-ОБРАЗНАЯ ЧУШЬ

РАЗОБЛАЧЕНИЕ МОШЕННИЧЕСТВА, ПРЕДВЗЯТОСТИ, НЕДОБРОСОВЕСТНОСТИ И ХАЙПА В НАУКЕ

Перевод с английского
Алёны Якименко



издательство **АСТ**

Москва

УДК 001:159.9
ББК 72:88.3
Р56

Издание осуществлено при поддержке “Книжных проектов Дмитрия Зимина”

Художественное оформление и макет Андрея Бондаренко

Ричи, Стюарт.
Р56 Наукообразная чушь. Разоблачение мошенничества, предвзятости, недобросовестности и хайпа в науке / Стюарт Ричи; пер. с англ. А. Якименко. — Москва : Издательство АСТ : CORPUS, 2024. — 432 с.

ISBN 978-5-17-136104-4

Чтобы определиться со своим мировоззрением и принимать взвешенные решения, влияющие на нашу жизнь, нам среди прочего необходимо сформировать свое отношение к высказываниям экспертов и результатам научных исследований. “Ученые врут и все скрывают” или “ученые доказали, значит, так и есть”? Между двумя этими крайностями лежит огромное поле неоднозначности, и книга Стюарта Ричи помогает нам сориентироваться на нем, найти свою позицию. Приводя потрясающие воображение примеры, Ричи описывает системные проблемы в публичном освещении научных работ и устройстве самой научной деятельности, позволяя критически переосмыслить все, что мы читали, смотрели и слушали о науке до сих пор. Однако надежды на прогресс он у нас вовсе не отнимает — напротив, мы много узнаем о том, что делается и что еще можно сделать, чтобы доверие к работе ученых крепло на более надежных и устойчивых основаниях.

УДК 001:159.9
ББК 72:88.3

ISBN 978-5-17-136104-4



Книжные проекты Дмитрия Зимина

Эта книга издана в рамках программы
“Книжные проекты Дмитрия Зимина”
и продолжает серию
“Библиотека фонда «Династия»”.
Дмитрий Борисович Зимин —
основатель компании “Вымпелком” (*Beeline*),
фонда некоммерческих программ “Династия”
и фонда “Московское время”.

Программа “Книжные проекты Дмитрия Зимина”
объединяет три проекта, хорошо знакомых
читательской аудитории:
издание научно-популярных
книг “Библиотека фонда «Династия»”,
издательское направление фонда “Московское время”
и премию в области русскоязычной
научно-популярной литературы
“Просветитель”.

Подробную информацию
о “Книжных проектах Дмитрия Зимина”
вы найдете на сайте
ZIMINBOOKPROJECTS.RU

Содержание



Посвящается Кэтрин

Вот *это* — научный факт. Реальных *доказательств* тому нет, но это научный факт.

Из сериала *Brass Eye*[1]

Предисловие

> …Все же уму человеческому постоянно свойственно заблуждение, что он более поддается положительным доводам, чем отрицательным…*
>
> Фрэнсис Бэкон *"Новый Органон" (1620)*[2]

В январе 2011 года мир узнал, что студенты обладают экстрасенсорными способностями.

Результаты из новой научной статьи произвели сенсацию: в серии лабораторных экспериментов с участием более чем тысячи человек было доказано существование экстрасенсорного предвидения — умения заглянуть в будущее с помощью сверхчувственного восприятия. Статью написал не какой-то неизвестный безумец, а ведущий профессор психологии Дэрил Бем из Корнеллского университета, входящего в Лигу плюща. И опубликована она была не абы где, а в одном из самых главных научных журналов по психологии, уважаемом и рецензируемом[3]. Казалось, наука официально одобрила феномен, до того считавшийся абсолютно невозможным.

Я тогда был аспирантом, занимался психологией в Эдинбургском университете. И прилежно ознакомился со статьей Бема. Вот в чем заключался один из экспериментов. Студенты смотрели на экран компьютера, где появлялись два изображения занавеса. Им сказали, что за одним из изображений есть еще картинка и нужно предположить, за каким именно, нажав на него. Поскольку никакой дополнительной информации предоставлено не было, студенты могли лишь гадать. Ко-

* Перевод С. Красильщикова.

гда они делали выбор, занавес исчезал — и они видели, правы оказались или нет. После тридцати шести повторений эксперимент завершался. Получилось нечто странноватое. Когда за занавесом прятался какой-нибудь нейтральный, скучный объект вроде стула, результат был почти идеально случайным: студенты делали правильный выбор в 49,8 % случаев, то есть вероятность угадать была, по сути, пятьдесят на пятьдесят. Однако — и в этом-то вся странность — когда спрятанная картинка была порнографической, студенты выбирали соответствующее изображение занавеса чуточку чаще, чем это предписывается простым угадыванием: в 53,1 % случаев, если быть точным. Разница превысила порог так называемой статистической значимости. В статье Бем предположил, что некое подсознательное, экстрасенсорное половое влечение самую малость подталкивало студентов к эротической картинке еще до того, как та появлялась на экране[4].

Некоторые другие эксперименты Бема были менее откровенными, но не менее обескураживающими. В одном из них на экране поочередно появлялись сорок не связанных между собой слов. Затем проводилась проверка памяти — студенты должны были напечатать столько слов, сколько им удалось запомнить. Далее компьютер случайным образом выбирал двадцать слов из списка и снова высвечивал их на экране. На этом эксперимент завершался. Бем сообщил, что во время теста на запоминание студенты чаще припоминали те двадцать слов, *которые вскоре снова увидят*, хоть и не могли знать — разве что благодаря некой сверхъестественной интуиции, — какие именно слова им покажут. Как если бы студент готовился к экзамену, сдавал его, затем занимался еще уже после сдачи — и вот эта постэкзаменационная зубрежка каким-то образом прокладывала себе путь назад во времени, улучшая отметку. Времени положено идти лишь в одном направлении, если, конечно, законы физики внезапно не утрачивают своей силы. Причинам положено предшествовать своим следствиям, а не наступать после них. Однако публи-

кация статьи Бема сделала эти странные результаты частью научной литературы.

Что важно, эксперименты Бема совершенно нехитрые, они не требуют ничего сложнее стационарного компьютера. Если Бем прав, любой исследователь мог получить доказательство существования паранормального, просто следуя его описанию того, как проводились эксперименты, — даже аспирант с примерно нулевыми ресурсами, каким был я. Именно это я и сделал: связался с двумя другими психологами, тоже сомневавшимися в объявленных результатах, Ричардом Вайзманом из Университета Хартфордшира и Крисом Френчем из Голдсмитского университета Лондона, и мы договорились трижды повторить эксперимент Бема со списком слов, по разу в каждом из трех наших солидных университетов. Несколько недель спустя, в течение которых мы набирали участников, ждали, пока они пройдут тест на запоминание, и смотрели, как на их лица наползает недоумевающее выражение, когда в конце мы объясняли, что именно проверяем, — итак, несколько недель спустя мы получили результаты. И они… не показали ничего. Наши студенты не были ясновидцами: слова, которые после теста снова высвечивались на экране, они припоминали не лучше и не хуже, чем остальные. Пожалуй, законы физики все-таки были вне опасности.

Мы надлежащим образом изложили свои результаты и отослали статью в тот же научный журнал, где была опубликована работа Бема, — *Journal of Personality and Social Psychology*. И почти моментально нам дали от ворот поворот. Редактор отклонил статью в течение нескольких дней, аргументировав это тем, что политика журнала — никогда не публиковать статьи, посвященные повторению каких-то предыдущих экспериментов, вне зависимости от того, воспроизвелись прежние результаты или нет[5].

Разве неправы мы были, почувствовав себя оскорбленными? Журнал опубликовал статью, в которой были сделаны чрезвычайно смелые заявления — заявления, не просто пред-

ставляющие интерес для психологов, а такие, что в корне изменили бы науку, окажись они правдивы. Результаты этой статьи попали в публичное пространство и получили широкую огласку в популярных медиа, в частности, Бема пригласили на вечернюю телепрограмму *The Colbert Report*, где ведущий употребил эффектное выражение "порно, путешествующее во времени"[6]. И однако же редакторы журнала отказались даже рассматривать возможность публикации статьи, рассказывающей о повторении исследования и ставящей сенсационные результаты под сомнение[7].

А тем временем развертывался и другой эпизод, тоже поднимающий тревожные вопросы о современном состоянии научной практики. Журнал *Science*, считающийся одним из самых престижных научных журналов в мире (уступающий разве что *Nature*), опубликовал статью Дидерика Стапела, социального психолога из Тилбургского университета в Нидерландах. В статье под заглавием "Преодолевая хаос" описывалось несколько исследований, проведенных в лаборатории и на улице, которые продемонстрировали, что люди сильнее подвержены предубеждениям — и расовым стереотипам — в более беспорядочной и загрязненной среде[8]. Эта статья и некоторые другие из десятков работ Стапела произвели фурор во всем мире. "Хаос способствует стереотипам" — написала новостная служба журнала *Nature*; "Где мусор — там расизм" — грянула *Sydney Morning Herald*[9]. Это был пример того типа исследований в области социальной психологии, что дают легкие для понимания результаты с "ясными практическими выводами", как написал сам Стапел, — в данном случае следовало "заблаговременно выявлять беспорядок в окружающей среде и немедленно вмешиваться"[10].

Проблема заключалась в том, что все это было неправдой. Некоторые коллеги Стапела насторожились, заметив, что результаты его экспериментов *слишком уж* идеальны. Мало того, обычно серьезные ученые настолько загружены, что доверяют черную работу вроде сбора данных своим студентам, однако

Стапел будто бы сам выходил и собирал все данные собственноручно. После того как его коллеги в сентябре 2011 года поделились своей озабоченностью с университетским руководством, Стапел был отстранен от работы. За этим последовали многочисленные разбирательства[11].

В автобиографии-исповеди, которую он потом написал, Стапел признался: вместо того чтобы собирать данные для своих исследований, он допоздна засиживался один в своем кабинете или за письменным столом у себя дома и вбивал в таблицу числа, необходимые для его воображаемых результатов, просто-напросто их выдумывая. "Я делал ужасные вещи, возможно даже отвратительные, — писал он. — Я подделывал научные данные и сочинял исследования, которых никогда не было. Работал я один, отчетливо осознавая, что делаю… И ничего не чувствовал: ни отвращения, ни стыда, ни сожалений"[12]. Поразительно, но его мошенничество было сложноорганизованным и проработанным до мелочей. "Я выдумывал целые школы, где якобы проводил свои исследования, учителей, с которыми якобы обсуждал эксперименты, лекции, которые якобы читал, уроки по социальным наукам, проведению которых якобы способствовал, подарки, которые якобы дарил людям в благодарность за их участие в проекте"[13].

Стапел описал, как распечатывал пустые бланки будто бы для того, чтобы раздать участникам исследований, показывал их коллегам и студентам, заявляя, что вот-вот запустит новый проект, а потом вдали от чужих глаз… выбрасывал всю эту макулатуру в бак для переработки. Это не могло больше продолжаться. Результаты инициированных разбирательств были однозначны — Стапела уволили вскоре после временного отстранения от работы. С тех пор не менее пятидесяти восьми его статей были отозваны — изъяты из научной летописи — из-за фальсификации данных.

Случаи с Бемом и Стапелом, когда уважаемые ученые публиковали на первый взгляд невозможные (Бем) или откровенно жульнические (Стапел) результаты, хорошо встрях-

нули дисциплину психологических исследований и вообще всю научную область. Как авторитетные научные журналы допустили эти работы до публикации? И скольким еще опубликованным ранее статьям нельзя верить? Как выяснилось, эти случаи послужили прекрасным примером гораздо более глубоких проблем в том, как мы занимаемся наукой.

В обоих случаях ключевой вопрос имел отношение к *воспроизведению результатов*. Научное открытие заслуживает того, чтобы его воспринимали всерьез, если это не что-то возникшее по случайности, или из-за неисправностей оборудования, или по милости ученого, мухлюющего или замалчивающего какие-то нестыковки. Это открытие должно действительно произойти. И если оно произошло, то я теоретически могу пойти и получить примерно те же результаты, что и вы. Во многом это и есть суть науки, именно это отличает ее от других способов познания мира: если какой-то результат не воспроизводится, трудно вообще назвать его научным.

Итак, тревогу вызвало не столько то, что эксперименты Бема были сомнительными, а Стапела — плодом его воображения (какие-то просчеты и ложные результаты неизбежны, они будут всегда[14], а стало быть, увы, всегда будут и мошенники), настоящая проблема заключалась в том, как научное сообщество повело себя в обеих ситуациях. Наша с коллегами статья, посвященная воспроизведению экспериментов Бема, была бесцеремонно отвергнута журналом, опубликовавшим исходную работу, а в случае Стапела почти никто даже никогда не *пытался* воспроизвести его результаты[15]. Иными словами, научное сообщество продемонстрировало, что согласно принимать на веру громкие заявления, звучащие в этих статьях, не проверяя, насколько на самом деле излагаемые результаты достоверны. А если воспроизводимость результатов никак не проверяется, откуда нам вообще знать, что они не просто случайны или сфальсифицированы?

Пожалуй, лучше всего отношение многих ученых к воспроизведению экспериментов выразил сам Бем в одном ин-

тервью через несколько лет после его печально известного исследования. “Я обеими руками за научную строгость, — сказал он, — но мне самому не хватает на нее терпения… Если бы вы ознакомились с предыдущими моими экспериментами, вы бы увидели, что все они служили своего рода риторическими приемами. Я собирал данные для того, чтобы изложить свою точку зрения. Я использовал данные как средство убеждения и, в сущности, никогда не переживал о том, воспроизведутся мои результаты или нет”[16].

Беспокоиться, воспроизведутся ли результаты, — отнюдь не что-то опциональное. В этом основной смысл науки, смысл, который должен явно обнаруживаться в системе рецензирования и журнальных публикаций, играющей роль бастиона на пути ошибочных открытий, ложно понятых экспериментов и сомнительных данных. Однако, как будет показано дальше, эта система серьезно поломана. Ученые получают важные результаты, но они расцениваются как не слишком интересные для публикации, так что их изменяют или прячут, искажая научную летопись и вредя нашей медицине, технологиям, методикам обучения и государственной политике. Огромные ресурсы, вливаемые в науку в надежде на полезную отдачу, растрачиваются на исследования, абсолютно не несущие новых знаний. Легко предотвратимые ошибки и ляпы регулярно прорывают линию Мажино системы рецензирования. Книги, сообщения средств массовой информации и наши головы наполняются “фактами”, которые неверны, преувеличены или заведомо ложны. И в самых худших случаях, особенно когда дело касается медицинской науки, гибнут люди.

В других книгах ученые выводятся как борцы с целой галереей аферистов, представителей псевдонауки: с креационистами, гомеопатами, плоскоземельцами, астрологами и иже с ними, которые своей деятельностью олицетворяют полное непонимание науки и совершают над ней надругательство — обычно неосознанно, иногда злонамеренно и всегда безответственно[17]. Моя же книга иная. В ней раскрывается глубин-

ное разложение самой науки: разложение, влияющее на среду, в которой проводятся исследования и публикуются их результаты. Наука — практика, которая должна отличаться строжайшим скептицизмом, самым острым рационализмом и трезвым эмпиризмом, — стала вместилищем ошеломляющего количества заблуждений, некомпетентности, лжи и самообмана. И при этом основная цель науки — пробиваться как можно ближе к истине — подрывается.

В первой части этой книги будет показано, что занятие наукой подразумевает куда больше, чем просто проведение экспериментов и проверку гипотез. Наука по своей сути *социальна*, ведь вы должны убеждать других людей — остальных ученых — в том, что обнаружили. А поскольку наука — это еще и *человеческая* сфера, мы понимаем, что любой ученый будет подвержен человеческим особенностям, таким как иррациональность, предвзятость, сбои внимания, внутригрупповой фаворитизм и откровенное жульничество для получения желаемого. Чтобы ученые могли убеждать друг друга, стараясь при этом преодолеть неотъемлемые ограничения человеческой природы, наука создала систему сдержек и противовесов, которая — в теории — отделяет научные зерна от плевел. Такой процесс тщательного изучения и проверки на достоверность, приводящий к предполагаемому золотому стандарту публикаций в рецензируемых научных журналах, описывается в первой главе. Однако во второй главе показывается, что этот процесс катастрофически нарушен: в различных областях науки существуют бесчисленные опубликованные результаты, которые не воспроизводятся и достоверность которых крайне сомнительна.

Затем, во второй части книги, мы зададимся вопросом, почему так происходит. Мы обнаружим, что наша публикационная система, вместо того чтобы устранять или преодолевать всяческие человеческие факторы, позволяет им накла-

дывать свой отпечаток на научную летопись — и именно потому, что *считает* себя объективной и беспристрастной. Возобладала своеобразная самоуверенность, странная самонадеянность, когда само существование системы рецензирования словно бы мешает нам распознавать ее недостатки. Казалось бы, рецензируемые статьи должны настолько приближаться к объективной фактографии устройства нашего мира, насколько это вообще возможно. Однако в ходе обсуждения многих десятков таких статей мы увидим, что нельзя надеяться, будто рецензирование гарантирует, что ученые будут честно (глава третья), непредвзято (глава четвертая), добросовестно (глава пятая) обращаться со своими результатами и трезво их оценивать (глава шестая).

В третьей части книги мы глубже погрузимся в научную практику. В седьмой главе будет показано: дело не только в том, что система не справляется со всеми обсуждаемыми видами злоупотреблений, — в действительности современное устройство научно-исследовательской деятельности *стимулирует* возникновение этих проблем, подталкивая ученых к зацикленности на авторитете, известности, финансировании и репутации в ущерб строгости и надежности результатов. Наконец, после того как мы диагностируем проблему, в восьмой главе будет описан ряд зачастую радикальных реформ научной практики, которые могли бы помочь переориентировать ее в направлении исходной цели — открывать факты о нашем мире.

Чтобы обосновать слабости научно-исследовательского процесса, по ходу повествования я буду рассказывать поучительные истории из самых разнообразных областей науки. Отчасти потому, что я психолог, преобладать будут примеры из психологии[18]. Но в этой книге будет много говориться о психологии не только из-за моего образования и опыта работы. Причина также в том, что после инцидентов с Бемом и Стапелом (в числе многих других) психологи начали заниматься глубоким самокритическим анализом. По-

жалуй, активнее, чем коллеги в любой другой области науки, мы, психологи, стали выявлять свои укоренившиеся недостатки и разрабатывать систематические подходы для борьбы с ними — подходы, которые многие другие дисциплины начинают заимствовать.

Первый шаг в починке нашей сломанной научной системы — научиться распознавать и исправлять те ошибки, что могут сбить ее с пути. А сделать это возможно только с помощью опять же науки. В книге я постоянно буду прибегать к *метанауке* — сравнительно новому типу научного исследования, в котором изучается само научное исследование. Если наука — это процесс выявления и устранения ошибок, то метанаука представляет собой тот же процесс, но нацеленный на внутренние проблемы.

Ошибки многому могут научить. Музыкант Тодд Рандгрен во вводном треке одного из своих альбомов предлагает слушателям сыграть в игру под названием “Студийные звуки”. Он описывает всевозможные огрехи, которые могут проявиться при записи музыки: гудение, шипение, щелчки, когда в микрофон пропевают слова, содержащие букву “п”, прерывистый монтаж и тому подобное. Рандгрен предлагает слушателям считывать такие шероховатости при прослушивании его песен в этом альбоме и любых других записей. И как понимание огрехов звукозаписывающих студий может по-новому раскрыть для вас процесс создания музыки, точно так же понимание того, как наука отклоняется от верного пути, может многое сказать о процессе, благодаря которому мы постигаем новое знание.

Узнав о серьезных проблемах с тем, как мы занимаемся наукой, вы придете в замешательство. Сколько любопытных результатов, о которых вы читали в новостях и научно-популярных книгах или узнавали из документальных фильмов (открытий, настолько вас восхитивших, что вы поделились ими

с друзьями, или заставивших вас переосмыслить устройство нашего мира), основаны на слабых исследованиях, которые нельзя воспроизвести? Сколько раз врач прописывал вам лекарство или метод лечения, сведения об эффективности которого зиждутся на ущербных данных? Сколько раз вы меняли свой рацион, покупательские привычки или какой-то другой аспект образа жизни, руководствуясь научным исследованием, только чтобы убедиться, что его результаты через несколько месяцев будут полностью опровергнуты в новом исследовании? Сколько раз государственные деятели устанавливали законы или вырабатывали политику, напрямую влияющую на жизнь людей, ссылаясь на научные результаты, не выдерживающие критики? Ответ на каждый из этих вопросов один: это происходило гораздо чаще, чем вам хотелось бы думать.

Наивно рассчитывать, что абсолютно каждое научное исследование будет истинным — неким отчетом о незыблемых фактах, который никогда не подвергнется пересмотру в будущих исследованиях. Мир для этого чересчур беспорядочен. Надеяться можно только на то, что наши научные исследования заслуживают доверия, то есть честно сообщают, что произошло в экспериментах. Если хваленая система рецензирования не способна оправдать это доверие, наука утрачивает одно из своих основных и наиболее ценных качеств, наряду со способностью делать то, что она делает лучше всего: коренным образом менять наш мир благодаря постепенному появлению и развитию новых знаний, технологий, методик и средств лечения.

Моя цель — восхвалять науку, а не хоронить ее. В этой книге я отнюдь не нападаю на саму науку или ее методы. Напротив, я защищаю научные методы и вообще научные принципы от того, как сейчас занимаются наукой. Беды, о которых мы будем говорить, так тревожат именно из-за важности науки: позволяя ей померкнуть, а прогрессу страшно забуксовать, мы подвергаем себя риску разрушить одно из величайших достижений нашего вида.

Однако нанесенный ущерб поправим. Теоретически — а возможно, и в практическом плане — наука все еще способна быть функциональной и надежной системой знаний, какой она нам и нужна. Пока мы будем обсуждать в этой книге череду научных провалов, стоит держать в голове позитивную мысль (дрожащий проблеск надежды и утешения, что вырывается из ящика Пандоры, наполненного мошенничеством, предвзятостью, недобросовестностью и хайпом, который мы вскроем в следующих главах): почти все описываемые проблемы были обнаружены *другими учеными*. Умные идеи метанауки, предложенные для борьбы с этими проблемами и наведенным беспорядком, в значительной мере зародились внутри научного сообщества. Пусть во многих областях он и погребен глубоко, но дух самокритики, вдыхающий жизнь в подлинную науку, еще теплится.

И хорошо, что так, поскольку, как мы вскоре увидим, в науке царит изрядный бардак.

ЧАСТЬ I
Как должно быть — и как на самом деле

Глава 1
Как работает наука

Для того, чтобы ум мог заниматься такими предметами мышления надлежащим образом, недостаточно уединения, но требуется компания наших ближних и беседа с ними*.

Дэвид Юм *“О написании эссе” (1777)*[1]

Наука — это социальный конструкт.

И прежде чем из-за этой фразы вы отшвырнете книгу, позвольте объяснить, что я имею в виду. Я употребляю это выражение не в том смысле, в каком его используют крайние релятивисты, постмодернисты, поборники антинауки и прочие, кто полагает, что реального мира не существует, что наука — лишь один не так чтобы особенный способ узнать об этом или даже что наука — это просто один из “мифов” среди многих других, в которые мы могли решить верить[2]. Наука вылечивала болезни, картировала мозг, прогнозировала климатические условия, расщепила атом — это лучший из имеющихся у нас методов для выяснения того, как работает вселенная и как подчинить ее нашей воле. Другими словами, наука — это лучший способ приближения к истине. Разумеется, мы никогда не достигнем конца пути — достаточно беглого взгляда на историю, чтобы увидеть, сколь высокомерно было бы утверждать, будто какие-либо факты неопровержимы и неизменны. И все же для постепенного продвижения к большему знанию о мире лучше научных методов нет ничего.

Однако одними этими методами нам прогресса не достичь. Недостаточно в одиночку сделать какое-то наблюдение

* Перевод О. Артемьевой. Цит. по: *Своеволие философии: собрание философских эссе*. М.: ЯСК, 2019.

в своей лаборатории, вы должны еще убедить других ученых в том, что открыли нечто реальное. И вот тут-то и начинается социальная часть. Философы уже долго обсуждают, насколько важно для ученых *показывать* коллегам-исследователям, как они пришли к своим выводам. Вот как это сформулировал Джон Стюарт Милль[3]:

> Даже при изучении природы, и здесь всегда возможны различные объяснения одних и тех же фактов, возможна теория геоцентрическая и теория гелиоцентрическая, возможна и теория флогистона, и теория кислорода, — и чтобы признать какую-нибудь из них истинной, надо доказать, что другая не есть истина, а пока это не доказано или пока мы не знаем, как это доказывается, то, признавая одну из них истинной, не знаем, значит, оснований мнения, которого держимся*.

Поэтому ученые работают в командах, путешествуют по миру с лекциями и докладами на конференциях, дискутируют друг с другом на семинарах, организуют научные общества для обсуждения результатов исследований и — что, пожалуй, самое важное — публикуют свои результаты в рецензируемых журналах. Эти социальные аспекты — не преимущества профессии и не просто товарищество. Они представляют собой научный процесс в действии — беспрерывную череду коллективных актов: тщательно изучить, поставить под сомнение, проверить, уточнить и выработать *консенсус*. Хоть это и кажется на первый взгляд парадоксальным, но именно субъективный процесс науки и порождает ее беспримерную степень объективности[4].

Вот в каком смысле наука является социальным конструктом. Всякое утверждение о нашем мире может быть названо научным знанием только после того, как оно пройдет эту об-

* Перевод А. Неведомского.

щественную проработку, которая призвана отсеивать ошибки и упущения и давать возможность другим ученым судить, достойно ли новое открытие звания надежного, достоверного и важного. Благодаря тому, что каждое открытие прогоняют через такую полосу препятствий, конечные продукты научного процесса — публикуемые рецензируемые статьи — наливаются порядочной силой для общества. Мы говорим: это *наука*, а не просто шаблонная фраза, или болтовня, или чье-то мнение.

Социальная природа науки вместе с тем имеет и слабые места. Поскольку ученые так сильно сосредоточены на попытках убедить своих коллег, что подразумевает прохождение результатов исследования через этап рецензирования и дальнейшие шаги вплоть до публикации, им очень просто позабыть о настоящей цели науки — приближать нас к истине. А поскольку ученые еще и человеческие существа, способы, которыми они стараются убеждать друг друга, не всегда в полной мере рациональны или беспристрастны[5]. Если мы не будем особенно внимательны, научный процесс может пропитаться несовершенствами самого человека.

Эта книга посвящена тому, как мы уделяли недостаточно внимания научному процессу. Тому, как мы в итоге очутились в ситуации, когда научная система не просто не учитывает наши человеческие слабости, но еще и усиливает их. В последние годы стало донельзя очевидно, что рецензирование отнюдь не гарантирует корректности и надежности, как ему положено, а публикационный процесс, который должен быть оплотом науки, превратился в ее ахиллесову пяту.

Однако чтобы понять, как же научная публикационная система так испортилась, сначала нам нужно узнать, как ей надлежит работать в отлаженном состоянии.

Представим, что вы хотите заняться наукой. Первым делом вам придется проштудировать научную литературу. Она представляет собой обширнейшую коллекцию журналов для специа-

листов, изданий, которые служат главными источниками нового научного знания. Идея периодических изданий, где ученые могли бы делиться результатами своей работы, восходит к 1665 году, когда Генри Ольденбург из Королевского общества Великобритании опубликовал первый номер журнала, приведем его полное заглавие, *Philosophical Transactions: Giving Some Account of the Present Undertakings, Studies, and Labours of the Ingenious in Many Considerable Parts of the World* ("Философские труды, дающие некоторый отчет о нынешних начинаниях, исследованиях и трудах гениальных людей во многих значительных частях мира")[6]. Идея состояла в том, чтобы эти гениальные ученые могли присылать письма с описанием своих деяний для ознакомления заинтересованных читателей. До того ученые либо трудились в одиночку в угоду богатым правителям или частным покровителям и организациям (и их научные изыскания часто воспринимались скорее как некие салонные фокусы, чем как попытка найти истину) и издавали отдельные книги, либо формировали кружки единомышленников, которые обменивались письмами. Из такого вот клуба по переписке и выросли организации вроде Королевского общества[7].

Первые номера журнала Ольденбурга больше напоминали своеобразный вестник, где описывались недавние эксперименты и открытия. Скажем, в первом выпуске первого тома натурфилософ и энциклопедист Роберт Гук рассказал о самом первом наблюдении предположительно Большого красного пятна на Юпитере. Заметка вся целиком была такой:

> Гениальный мистер Гук несколькими месяцами ранее поставил своего друга в известность о том, что несколько дней назад, как он сказал (а именно 9 мая 1664 года, около девяти часов вечера), он наблюдал в великолепный двенадцатифутовый телескоп маленькое пятно на наибольшем из трех более темных поясов Юпитера и что, время от времени поглядывая на него, обнаружил, что за два часа оно переместилось с востока на запад примерно на половину диаметра Юпитера[8].

Журнал существует и по сей день, только под несколько более удобоваримым названием *Philosophical Transactions of the Royal Society* ("Философские труды Королевского общества")[9]. Время шло, и на смену коротким новостным заметкам пришли статьи подлиннее, содержащие подробные описания экспериментов и исследований. Сейчас это издание входит в глобальную экосистему из более чем тридцати тысяч журналов, от самых общих (как высокопрестижные *Nature* и *Science*, которые публикуют результаты самых значительных исследований со всего мира, относящихся к любой научной области) до узкоспециальных (вроде *American Journal of Potato Research*, интересующегося статьями исключительно на картофельную тему)[10]. Некоторые журналы издаются научными обществами, как — до сих пор — *Philosophical Transactions*, но большинство принадлежат коммерческим компаниям, таким как *Elsevier*, *Wiley* и *Springer Nature*[11]. Благодаря недавнему прогрессу все научные журналы были переведены в электронный вид, так что любой человек, который может заплатить взносы, взимаемые издателем с подписчиков, либо препоручить это библиотеке своего университета, овладевает мировым научным знанием, находящимся теперь всегда у него под рукой[12].

После прочтения журналов, относящихся к вашей области интересов, вы можете определиться с предметом исследования. Например, есть научная теория, в рамках которой делается некое предсказание — формулируется гипотеза, а вы можете каким-нибудь хитрым способом ее проверить. Или вы знаете, что в имеющихся у ученых знаниях есть пробел, а вы понимаете, как его заполнить. Или в порыве вдохновения вы придумали эксперимент, который позволит открыть нечто совершенно новое. Правда, прежде чем приняться за воплощение своей идеи, обычно требуется раздобыть денег на финансирование исследования: чтобы, к примеру, купить оборудование или материалы, привлечь участников или оплатить труд ученых, нанятых для сбора данных. Если вы, скажем, не фармацевтическая компания, которая может позволить себе дер-

жать собственные лаборатории, основной способ получить жизненно важное финансирование — это подать заявку на грант. Гранты предоставляются правительством страны, бизнес-компаниями, дотационными фондами, некоммерческими организациями, благотворительными фондами или даже просто состоятельными физическими лицами. Вы можете подать заявку в Национальные институты здравоохранения (NIH) или в Национальный научный фонд (это организации в США, финансируемые из государственного бюджета) либо в благотворительный фонд, поддерживающий науку, например *Wellcome Trust* или Фонд Билла и Мелинды Гейтс[13].

Финансирование отнюдь не гарантировано, и любой ученый подтвердит, что один из самых выматывающих аспектов научной работы — это поиск источников финансирования для своих новых исследований, а вероятность получить отказ жестоко высока. Вечные попытки найти деньги оказывают важное косвенное влияние на саму науку, и мы еще к этому вернемся. Однако сейчас давайте представим, что вам повезло получить грант. Тогда вы приступаете к работе. На этапе сбора данных может потребоваться сталкивать между собой частицы в подземном коллайдере, искать окаменелые остатки в арктической зоне Канады, воссоздавать в чашке Петри точные условия для роста бактерий, организовывать посещение лаборатории и заполнение опросников для сотен людей, прогонять сложную математическую модель на компьютере. Этот этап может занять дни, месяцы, десятилетия.

Когда данные уже получены, вы, как это обычно бывает, располагаете набором чисел, который можете — сами или с помощью более подкованного в математике коллеги — проанализировать различными статистическими методами (это еще одно опасное место, куда мы позже вернемся). Затем вы должны все это описать в форме научной статьи. Типичная статья начинается с раздела “Введение”, где вы кратко излагаете, что уже известно по этой теме и что добавляет ваше исследование. Далее следует раздел “Методы”, где вы описы-

ваете, что конкретно делалось в ходе исследования, причем настолько подробно, чтобы любой человек мог — теоретически — снова провести точно такой же эксперимент. Потом вы переходите к разделу “Результаты”, где в подтверждение своих открытий демонстрируете числа, таблицы, графики и статистический анализ. А венчает статью раздел “Обсуждение”, где вы свободно строите спекулятивные теории — то есть вдумчиво и на основе фактов рассуждаете — о том, что все это значит. Статью вы предваряете разделом “Аннотация” — кратким, обычно около ста пятидесяти слов, описанием исследования и его результатов. Аннотация всегда доступна для прочтения кем угодно, даже если доступ к самой статье ограничен подпиской на журнал, поэтому имеет смысл подходить к написанию аннотации со всем тщанием, чтобы полученные вами результаты звучали убедительно. Научные статьи бывают самой разной длины и объема, и иногда описанная очередность разделов нарушается, но в целом ваша статья будет выстроена по этому принципу[14].

Когда статья написана, вы вступаете в мир научных журналов и борьбы за публикацию. Еще недавно для того, чтобы подать статью в журнал, нужно было напечатать несколько экземпляров и отправить по почте редактору, но теперь все осуществляется через интернет (хотя у многих журналов до сих пор столь допотопные и ненадежные веб-страницы, что вы с таким же успехом могли бы отправить им статью с почтовым голубем). Редактор журнала — зачастую это солидный ученый — прочитает вашу статью (или, что не исключено, будем уж честны, только аннотацию) и решит, достойна ли она того, чтобы ее опубликовать. Бóльшая часть журналов, особенно самые престижные, кичатся своей исключительностью, а следовательно, и высоким порогом попадания к ним на страницы (*Science*, например, принимает менее 7% статей, поданных на рассмотрение), так что большинство статей на этой стадии будет отклонено без направления на рецензирование[15]. Это первый этап контроля качества: редактор от-

деляет статьи, соответствующие темам журнала и перспективные с точки зрения их научного значения и качества, от тех, которые не стоят дальнейшего рассмотрения. Для доли статей, что приглянулись редактору, наступает этап рецензирования. Редактор находит двух-трех ученых, являющихся экспертами в вашей области исследования, и спрашивает, не согласятся ли они оценить вашу статью. Они могут отказаться, сославшись на занятость, так что редактор продолжит идти по списку потенциальных рецензентов до тех пор, пока несколько из них не дадут согласие. И вот тут наступает этап нервного ожидания: одобрят ли они вашу работу.

Большинство людей, включая ученых, полагают, что рецензирование всегда было ключевым моментом для научной публикации, однако его история сложнее. Хотя в XVII веке Королевское общество стремилось спрашивать кого-то из своих членов, считают ли они некую статью достаточно интересной для публикации в *Philosophical Transactions*, по крайней мере до 1831 года не требовалось давать письменную оценку каждой работе[16]. И даже тогда формальная система рецензирования, какой мы ее знаем сегодня, не была универсальной, она стала таковой лишь в XX веке, да и то далеко не сразу (как можно понять из письма Альберта Эйнштейна, которое он послал в 1936 году редакторам журнала *Physical Review*, раздраженно заявляя, что снимает свою статью с рассмотрения в их журнале, поскольку они посмели отправить ее другому физику для оценки)[17]. Только в 1970-х годах уже все журналы начали применять современную практику отправки поданных на рассмотрение статей независимым экспертам на рецензирование, присвоив им роль фильтра, какую они и играют сегодня[18].

Обычно рецензенты выполняют свою работу анонимно, что одновременно и благословение, и проклятие. Благословение — потому что анонимность позволяет им без обиняков высказывать свою точку зрения, не волнуясь насчет ответной реакции ученых, чью работу они критикуют (зеленый научный сотрудник может действительно откровенно высказаться

о недостатках работы именитого профессора), а проклятие — поскольку анонимность, что уж тут, позволяет им без обиняков высказывать свою точку зрения, не волнуясь насчет ответной реакции ученых, чью работу они критикуют. Вот несколько выдержек из настоящих рецензий:

- Некоторые статьи читать одно удовольствие. Эта явно не из таких.
- Результаты слабые, какая-то размазня.
- Боюсь, эта статья послужит не столько прогрессу в данной области, сколько ее окончательному упадку.
- В рукописи делается три утверждения: первое известно уже годы, второе — десятки лет, третье — столетия.
- Когда вы писали это предложение, у вас не случилось припадка? Я вот бился в конвульсиях, пока читал его[19].

Если оценки рецензентов выглядят подобным образом, редактор, вероятно, отклонит вашу статью. Тогда вам захочется либо отступиться, либо начать весь процесс заново, подав статью на рассмотрение в другой журнал, а если и туда не примут, еще в один, а если снова не примут, в следующий и так далее — нередко статьи проходят через полдюжины или больше журналов, обычно со все более низким рейтингом, прежде чем их примут к публикации. Если же рецензенты впечатлятся больше, вы получите шанс доработать свою статью, учтя их замечания, например заново проанализировав данные, проведя дополнительные эксперименты или переписав какие-то разделы, и снова отправить ее редактору. Такая пересылка статьи туда-сюда в ходе ее доработки может иметь несколько итераций и часто длится месяцами. В конце концов, если рецензенты оказываются удовлетворены, редактор дает добро — и статья публикуется. Если журнал до сих пор выходит и в бумажной форме, вам доведется увидеть свою драгоценную статью в напечатанном виде, а иначе придется довольствоваться тем, чтобы с трепетом ждать ее появления

на официальном сайте журнала. Вот и все. Вы оставили свой след в научной литературе, и у вас теперь есть публикация, которую вы можете добавить в свое резюме и на которую другие ученые могут ссылаться. Поздравляю, до конца дня можете отдыхать.

Описание всего процесса, данное мной выше, разумеется, слишком краткое и общее, но по существу такой процесс в той или иной форме присущ каждой научной области. Мы должны спросить себя: после того, как статья вышла из-под катка рецензирования, в окончательном своем виде она все еще дает правдивое представление о том, что делалось в исследовании? Мы вернемся к этому вопросу позже. А пока нужно обсудить кое-что другое. Где гарантии, что участники описанного процесса — исследователь, подающий статью на рассмотрение, редактор журнала, рецензенты, оценивающие работу, — все проявляют честность и беспристрастность, каких требует истинная наука? Нет такого закона, который требовал бы, чтобы все руководствовались объективностью и рациональностью, когда имеют дело с наукой, поэтому требуется коллективный моральный устав, свод правил, направляющих поведение ученых[20]. Самая известная попытка записать эти неписаные правила была предпринята социологом Робертом Мертоном.

В 1942 году Мертон выдвинул четыре научных ценности, сегодня известных как “мертоновские нормы”. Называются они не слишком оригинально, зато служат правильными ориентирами для ученых. Первая — *универсализм*. Научное знание есть научное знание независимо от того, кто его формулирует, — при условии, что методы обретения этого знания надежны. Раса, пол, возраст, гендер, сексуальная ориентация, уровень доходов, социальное происхождение, национальность, известность или любая другая характеристика ученого не должна отражаться на том, как оцениваются его фактологические утверждения. Еще нельзя судить о чьем-либо исследовании на основании того, приятен или неприятен этот человек (к облегчению некоторых из моих менее приветли-

вых коллег). Вторая похожая норма — *бескорыстность*. Ученые не должны заниматься наукой из-за денег, по политическим или идеологическим причинам, ради славы (или репутации своего университета, страны и чего-либо еще) или чтобы потешить свое самолюбие. Они должны заниматься наукой во имя лучшего понимания вселенной, делая открытия и созидая, — и точка[21]. Как написал однажды Чарльз Дарвин, у ученого "не должно быть ни желаний, ни привязанностей — лишь каменное сердце"[22].

Оставшиеся две нормы напоминают нам о социальной природе науки. Третья — *коллективизм*. Она предписывает ученым делиться знаниями друг с другом[23]. Этот принцип лежит в основе самой идеи публикации ваших результатов в журнале, с тем чтобы о них узнали другие, — мы занимаемся наукой сообща, мы должны знать детали работы других ученых, чтобы иметь возможность проанализировать и развить ее[24]. И наконец, последняя норма: *организованный скептицизм*. Ничто не свято, и научное утверждение никогда не должно приниматься на веру. Нам следует воздерживаться от вынесения оценки любому открытию, пока мы тщательнейшим образом не проверили все данные и методы исследования. Самое очевидное воплощение нормы организованного скептицизма — рецензирование как таковое.

В теории все выглядит неплохо: придерживаясь четырех мертоновских норм, мы должны прийти к научной литературе, которой можно доверять, — к плечам гигантов, стоя на которых, по знаменитому выражению Ньютона, мы видим дальше. Разумеется, эти гиганты часто ошибались, взять хотя бы два примера Джона Стюарта Милля, процитированных выше: когда-то мы верили, что Солнце обращается вокруг Земли и что горючие вещества наполнены особой субстанцией — флогистоном, — высвобождающейся при их горении[25]. Но в итоге мы отбросили эти теории, когда получили новые данные.

И действительно, для ученого готовность изменить свое мнение — это достоинство. Биолог Ричард Докинз описал случай, которому стал свидетелем лично, это произошло с “уважаемым престарелым мужем” с факультета зоологии Оксфордского университета[26].

> Многие годы он страстно верил и учил студентов, что аппарат Гольджи (микроскопическая внутриклеточная структура) на самом деле не существует, что это — погрешность наблюдения, иллюзия. Каждый понедельник после обеда на факультете было заведено слушать научный доклад какого-нибудь заезжего лектора. В один из понедельников лектором оказался американский специалист по биологии клетки, представивший неотразимо убедительные свидетельства реальности аппарата Гольджи. В конце его выступления старик пробрался к подиуму и, пожимая американцу руку, с чувством провозгласил: “Дорогой коллега, позвольте выразить вам мою благодарность. Все эти пятнадцать лет я заблуждался”. Мы тогда аплодировали до боли в ладонях. <…> Не каждый ученый в реальной жизни [смог бы произнести такое]. Но для всех ученых подобные поступки являются эталоном — в отличие, скажем, от политиков, которые могли бы счесть старика беспринципным. У меня до сих пор комок к горлу подступает при воспоминании о том вечере*.

Вот что имеется в виду, когда говорят, что наука — это процесс “самокорректирующийся”. В конце концов, даже если это занимает много лет или десятилетий, старые неверные идеи ниспровергаются благодаря новым данным (а иногда, по довольно мрачному выражению физика Макса Планка, благодаря тому, что непреклонные поборники этих старых идей вымирают, оставляя науку следующему поколению[27]). И опять: так дело обстоит в теории. На практике же публика-

* Перевод Н. Смелковой.

ционная система, описанная выше, опирается на мертоновские нормы неуклюже, во многом препятствуя процессу самокорректировки. Характерные особенности этого противоречия — между конкуренцией за гранты и шумихой вокруг престижных публикаций с одной стороны и открытой, бесстрастной, скептической оценкой науки с другой — будут становиться все очевиднее в последующих главах.

А пока отметьте, что именно заставило престарелого мужа из рассказа Докинза изменить свое мнение — “неотразимо убедительные свидетельства”. Едва ли имеет смысл пытаться скорректировать или уточнить наши научные теории в соответствии с новыми данными, если сами данные неубедительны, а еще хуже — если они вообще неверны. Это возвращает нас к идее, которую мы обсуждали в предисловии: чтобы результаты заслуживали нашего доверия, они должны быть воспроизводимы. Вот как это сформулировал философ науки сэр Карл Поппер[28]:

> Только тогда, когда некоторые события повторяются в соответствии с некоторыми правилами и регулярностями (как в случае воспроизводимых экспериментов), наши наблюдения в принципе могут быть проверены каждым человеком. Даже наши собственные наблюдения мы не принимаем всерьез и не приписываем им статус научных наблюдений до тех пор, пока не повторим и тем самым не проверим их. Только в результате подобных повторений мы можем убедить себя в том, что имеем дело не с простым “совпадением”…*

Эта идея не то чтобы революционна — или была нова для Поппера, написавшего эти слова в 1950-х годах. Вернувшись в XVII век, когда журнал *Philosophical Transactions* только зародился, мы обнаружим, что сооснователь Королевского общества, химик Роберт Бойль, делал все возможное и невозмож-

* Перевод В. Брюшинкина.

ное, чтобы убедиться в воспроизводимости своих результатов. Он многократно демонстрировал эксперименты по установлению различных свойств воздуха и вакуума с использованием своего знаменитого воздушного насоса перед толпами зевак, а после заставлял их подписывать свидетельские показания о том, что они действительно наблюдали демонстрировавшийся процесс[29]. Он следил, чтобы его труды были написаны достаточно обстоятельно, дабы "человек, коему я их адресовал, сумел без ошибок и особых помех повторить столь необычные эксперименты"[30]. И хотя соорудить сложный аппарат было весьма непросто, он призывал других натурфилософов повторить его эксперименты с воздушным насосом в других частях Англии и Европы и оказывал им в том содействие[31].

Таким образом, воспроизведение экспериментов долгое время было ключевым ингредиентом рецепта, по которому должна была работать наука (и кстати, это еще один ее социальный аспект: результаты воспринимаются всерьез только после того, как окажутся подтверждены несколькими исследователями). Но с тех пор, где-то между Бойлем и современным академическим сообществом, громадное число ученых позабыли о важности повторений. В столкновении наших мертоновских идеалов с реалиями научной публикационной системы — не говоря уже о реалиях человеческой природы — идеалы оказались более хрупкими, оставив нас с научной литературой, полной ненадежных, недостоверных и невоспроизводимых результатов, которые зачастую больше путают, чем обогащают новым знанием.

В следующей главе мы узнаем, *насколько* научная литература запружена ненадежными, недостоверными и невоспроизводимыми результатами.

Глава 2
Кризис воспроизводимости

Взовьется на дыбы желанье власти
И валится, споткнувшись, в тот же миг*.

Уильям Шекспир *"Макбет", акт 1, сцена 7*

"Опубликовано" и "достоверно" — не синонимы.

Райан Носек, Джеффри Спайс и Мэтт Мотел[1]

Несомненно, самая популярная книга по психологии за минувшее десятилетие — это "Думай медленно… решай быстро" Даниэля Канемана. Не так много существует руководств по человеческому разуму, превосходящих канемановское. Он получил Нобелевскую премию по экономике 2002 года за свои исследования человеческой (ир)рациональности и опубликовал результаты десятков искусных экспериментов, демонстрирующих ограничения нашей способности мыслить. Книга "Думай медленно… решай быстро" стала сенсацией, суммарные продажи исчислялись миллионами экземпляров, и она до сих пор недурно расходится. И на то есть причины: это живо и доступно написанный путеводитель по всем ошибкам и искажениям человеческого мышления[2]. Помимо многих других тем Канеман осветил работы по "праймингу", как называют это явление психологи. Некоторые примеры прайминга (эффекта предшествования) имеют отношение к языку. Например, установлено, что если я показываю вам на экране компьютера набор слов по одному и прошу нажимать на клавишу всякий раз, как появляется слово "ложка", то вы будете реагировать чуточку быстрее, если перед тем высвечивалось слово "вилка"

* Перевод М. Лозинского.

(или какой-то еще из столовых приборов), чем когда предшествующим словом было “дерево” (или что-то еще не относящееся к принадлежностям для приема пищи). Восприятие слова “вилка” психологически “настраивает” вас быстрее реагировать на близкое по смыслу слово[3].

Канеман, однако, описал нечто более неожиданное. Он осветил результаты исследования по социальной психологии, демонстрирующие, что формирование установок, связанных с определенными *понятиями*, — обычно неосознанное — может ощутимо влиять на наше поведение. Один пример известен как “эффект леди Макбет”. В 2006 году в журнале *Science* вышла статья о том, что, когда исследователи просили участников эксперимента переписать текст об аморальных поступках, те потом с большей вероятностью хотели купить мыло, а просьба вспомнить что-то неприглядное из собственных деяний заставляла их чаще брать при выходе из лаборатории дезинфицирующую салфетку (“Проклятое пятно!”). Здесь воздействие оказывалось без *словесных* установок: получалось, что мозг работает куда более связно и целостно, чем мы полагали, что между понятиями и концепциями, вроде бы связанными очень условно, образуются прочные сшивки. В данном случае это казалось свидетельством некоего глубинного наложения понятий морали и чистоты. Авторы статьи утверждали, что эти результаты могут даже объяснить, почему омовение рук является частью столь многих религиозных ритуалов по всему миру[4].

Канеман также сделал обзор исследований “денежного прайминга”. В другой статье, вышедшей в *Science* в том же 2006 году, рассказывалось, как социальные психологи обнаружили, что, ненавязчиво напоминая людям о деньгах — скажем, сажая их за стол, где по случайности стоит компьютер с плавающими банкнотами на экранной заставке, — можно побудить их чувствовать и вести себя так, словно они более независимы, и меньше беспокоиться о других[5]. Авторы писали, что участники исследования, подвергшись денежному прай-

мингу, предпочитали “играть в одиночку, работать в одиночку и держать бóльшую физическую дистанцию между собой и новым окружением”[6]. И действительно, в ответ на просьбу расставить в комнате стулья для личной беседы с незнакомым человеком участники, подвергшиеся денежному праймингу, ставили стулья почти на тридцать сантиметров дальше один от другого по сравнению с участниками эксперимента, которые видели пустой экран компьютера. Нехилый эффект для простой экранной заставки, вероятно, подумали вы. И такой сценарий повторялся в большинстве значимых исследований прайминга: очень тонкие установки вызывали заметные изменения в поведении людей.

Канеман заключил, что подобные исследования прайминга “ставят под угрозу наше восприятие себя как сознательных и независимых творцов своих суждений и выборов”*. Он не сомневался в их надежности. И писал: “Не доверять этим данным нельзя. Они не выдумки и не статистические отклонения. Вам придется принять основные выводы исследований за истину. А еще важнее — принять их истинность применительно к *вам*”[7].

Но Канеман напрасно столь безоговорочно доверял этим эффектам предшествования, хоть они и были опубликованы в одном из самых уважаемых научных журналов. Оказывается, наряду с разоблачением мошенничества Дидерика Стапела и публикацией “сверхъестественных” результатов Дэрила Бема именно исследование, посвященное праймингу, — или скорее неудачная попытка его повторить — стало еще одним из исходных стимулов, подхлестнувших изучение явления, известного сегодня как “кризис воспроизводимости”[8].

В том исследовании прайминга участников просили найти в списке разрозненных слов одно лишнее, так чтобы остальные слова можно было сложить в связное предложение. У одной половины участников лишние слова были слу-

* Перевод “Школы перевода Баканова”.

чайными и нейтральными, а у другой — имеющими отношение к пожилым людям, например "старый", "седой", "мудрый", "вяжет" и "Флорида" (Флорида известна в Америке как штат, где живет много пенсионеров). Выполнив задание, участники исследования могли уходить — но в тайне от них экспериментаторы замеряли, как быстро те шли по коридору к выходу из здания. Демонстрируя опять-таки ментальную связь между идеями и действиями, те участники, кого подвергли праймингу связанными со старостью словами, *уходили из лаборатории медленнее* по сравнению с контрольной группой[9].

Опубликованная в 1996 году, статья об этом эксперименте с тех пор была процитирована другими исследователями более пяти тысяч раз и вошла во все учебники по психологии — помню, как сам студентом изучал ее[10]. Однако в 2012 году независимая группа попыталась провести точно такой же эксперимент, только с большей выборкой и совершеннее с технической точки зрения, — и не выявила никакой разницы в скорости ухода участников. Исследователи предположили, что в исходной работе получились такие результаты, поскольку сотрудники лаборатории, замерявшие время по секундомерам, знали, кто из участников как должен был бы себя вести, и это, вероятно, отражалось на замеряемых величинах. Измерение скорости ухода с помощью инфракрасных лучей, как было сделано в исследовании-повторении, свело к нулю предполагаемый эффект прайминга[11]. В течение нескольких лет другие лаборатории пытались воспроизвести как "эффект леди Макбет", так и эффект денежного прайминга, тоже с гораздо более внушительной и репрезентативной выборкой[12]. Эти попытки также очевидным образом провалились. Нет причин думать, что разнообразные результаты по праймингу были, процитируем Канемана, "выдумками", — нужно исходить из предположения, что получены они были добросовестным образом. Но вот "статистические отклонения"? Пожалуй, именно они.

Другие исследования эффекта предшествования были не лучше. В одном утверждалось, что участники, подвергшиеся праймингу "расстоянием" — их попросили нарисовать на листе миллиметровки две удаленные друг от друга точки, — чаще чувствовали себя дистанцированными от друзей и родных; попытка повторить это исследование провалилась в 2012 году[13]. В другом исследовании утверждалось, что когда примеры моральных дилемм распечатывались на листах с окантовкой в шахматную клетку, то участники эксперимента выносили более поляризованные суждения, поскольку этот узор заставлял их думать о концепции "белое и черное"; попытка повторить это исследование провалилась в 2018 году[14]. Сходное направление исследований, согласно которому возможно сделать людей более категоричными, подвергнув их праймингу "отвращением", было поставлено под сомнение в обзоре 2015 года[15].

Надо отдать Канеману должное — позже он признал, что был неправ, переоценив научную достоверность эффектов прайминга. "Экспериментальные доказательства идей, представленных мною в той главе, были значительно слабее, чем я полагал, когда писал ее, — сказал он через шесть лет после выхода книги "Думай медленно… решай быстро". — Это было попросту ошибкой: все, что я должен был знать для того, чтобы сдерживать свой энтузиазм, я знал… но не обдумал"[16]. Однако вред уже был нанесен: нобелевский лауреат объявил миллионам людей, что "нельзя не доверять" этим исследованиям.

Прайминг — не единственный психологический эффект, о котором узнали миллионы. Гарвардский психолог Эми Кадди прославилась в 2012 году, после того как выступила на конференции TED, восхваляя "позы силы". Непосредственно перед тем, как вы окажетесь в стрессовой ситуации, скажем на собеседовании, вам нужно потратить две минуты, советовала она, чтобы в каком-нибудь укромном уголке (вроде туалетной кабинки) встать в открытую, экс-

пансивную позу, например широко расставив ноги и уперев руки в бедра. Эта властная поза стимулирует вас психологически и гормонально. В эксперименте, проведенном Кадди и ее коллегами в 2010 году, обнаружилось, что по сравнению с людьми, которым предписано было сидеть скрестив руки или ссутулившись, те, кому велели принять позу силы, не только чувствовали себя более властными, но и шли на больший риск в азартной игре и имели повышенный уровень тестостерона и пониженный — кортизола, гормона стресса[17].

Идея Кадди, будто люди, на две минуты принимающие позу силы, могут “существенно менять исход жизненных ситуаций”, вызвала глубокий отклик: ее выступление на конференции TED стало вторым по количеству просмотров из всех — его посмотрели больше семидесяти трех с половиной миллионов раз[18]. Далее, в 2015 году, вышла книга Кадди по саморазвитию под названием “Присутствие [духа]”, бестселлер по версии газеты *The New York Times*, и издатель сообщал нам, что там представлена “захватывающая наука”, способная “освободить [нас] от страха в напряженные моменты жизни”[19]. Консервативная партия Великобритании, похоже, прониклась идеей Кадди, поскольку в тот самый год появилась череда фотографий, на которых представители этой партии на разных конференциях и выступлениях принимали позы с широко расставленными ногами, что вызвало немало насмешек[20]. В том же 2015 году другая команда ученых попыталась воспроизвести эффекты поз силы. И хотя те, кто принимал такие позы, действительно сообщали о большей уверенности в себе, исследование, увы, “не подтвердило влияние поз силы на тестостерон, кортизол и финансовый риск”[21].

Кризис воспроизводимости направил прожектор критики также и на более ранние работы по психологии — со сходными тревожными результатами. Вероятно, самое известное исследование по психологии за всю историю — это стэнфордский тюремный эксперимент 1971 года, когда пси-

холог Филип Зимбардо разделил группу молодых мужчин на "охранников" и "заключенных" и велел им неделю оставаться в импровизированной тюрьме в подвале факультета психологии Стэнфордского университета. Настораживающе быстро, по словам Зимбардо, "охранники" принялись наказывать "заключенных", мучая их столь садистски, что Зимбардо пришлось досрочно прекратить эксперимент[22]. Наряду с исследованиями подчинения, проведенными в 1960-х годах Стэнли Милгрэмом, в которых обнаружилось, что многие участники готовы воздействовать сильными электрическими разрядами на незадачливых "учеников" (удары током и ученики были ненастоящими, но участники об этом не знали), эксперимент Зимбардо приводится как одно из основных доказательств власти ситуации над человеческим поведением[23]. Поставьте, как говорится, хорошего человека в плохую ситуацию — и дела могут очень быстро пойти очень скверно. О стэнфордском тюремном эксперименте рассказывают фактически каждому студенту на планете, изучающему психологию, и Зимбардо благодаря ему стал одним из самых известных и уважаемых современных психологов. Он использовал результаты своего эксперимента, чтобы, например, выступать в качестве свидетеля-эксперта на суде над американскими военными, служившими охранниками в иракской тюрьме Абу-Грейб. Зимбардо утверждал, что ситуация, в которой оказались охранники, и роли, которые их заставили на себя взять, — вот причины их шокирующих издевательств и пыток над заключенными[24].

Хотя выводы из стэнфордского тюремного эксперимента всегда были противоречивы, лишь недавно мы начали понимать, насколько низкокачественным было то исследование[25]. В 2019 году ученый и кинорежиссер Тибо Ле Тексье опубликовал статью под названием "Разоблачение стэнфордского тюремного эксперимента". Он представил доселе неизвестную расшифровку видеозаписей, на которых Зимбардо вмешивается непосредственно в эксперимент, разда-

вая своим "охранникам" очень точные инструкции, как себя вести, — вплоть до того, что предлагает конкретные способы обесчеловечивания заключенных, например отказывать им в использовании туалетов[26]. Очевидно, столь основательно срежиссированная постановка эксперимента была далека от естественного примера того, что происходит, когда обычные люди оказываются в специфических социальных ролях. Как бы то ни было, несмотря на колоссальное внимание, которое долгие годы привлекал к себе стэнфордский тюремный эксперимент, его "результаты" с научной точки зрения не имеют смысла[27].

Как вы, наверное, догадались, психологов напугала совокупность неудавшихся попыток повторить эксперименты (как в исследованиях прайминга) и странных результатов (вроде паранормальных открытий Бема) наряду с разоблачением ложных представлений (как в эксперименте Зимбардо) и мошенничества (поддельные данные Стапела). Скольким же исследованиям в области психологии, недоумевали они, можно доверять? Чтобы получить представление о том, насколько дела плохи, они начали объединяться в команды для проведения крупномасштабных повторов значимых исследований в разных лабораториях. Самым заметным стало крупное сообщество ученых, отобравшее сто работ из трех топовых журналов по психологии и попытавшееся их воспроизвести. Читать о результатах, опубликованных в 2015 году в *Science*, было горько: в конечном счете лишь 39 % работ были признаны успешно воспроизведенными[28]. В другом подобном предприятии 2018 года ученые пытались повторить двадцать одно исследование по социальным наукам из двух самых авторитетных многопрофильных журналов в мире — *Nature* и *Science*. На сей раз воспроизвелось 62 %[29]. В последующих масштабных повторах исследований, касающихся разнообразных психологических феноменов, воспроизвелось 77, 54 и 38 % результатов[30]. Почти все повторы, даже успешные, продемонстрировали, что в исходных статьях эффекты были преуве-

личены. В целом кризис воспроизводимости легким движением руки стер, похоже, около половины всех исследований по психологии[31].

Возможно, все не столь уж плохо — по двум причинам. Во-первых, следует ожидать, что некоторые результаты, на самом деле надежные, иногда не получается воспроизвести просто по невезению[32]. Во-вторых, какие-то повторы могли провалиться из-за того, что их проводили с небольшими изменениями в методологии по сравнению с исходным исследованием (правда, если результат настолько нестабилен, что исчезает при малейших изменениях в постановке эксперимента, то возникает вопрос, имеет ли он вообще какой-то смысл и применение)[33]. По этим причинам иногда трудно понять, является результат “воспроизводимым” или нет, на основании только одной-двух попыток его повторить. Кроме того, доля воспроизводимых исследований для разных областей психологии, похоже, отличается: например, в статье 2015 года, вышедшей в *Science*, когнитивная психология (изучение памяти, восприятия, языка и так далее) проявила себя лучше, чем социальная (к коей относятся всевозможные исследования прайминга, обсуждавшиеся выше)[34].

Однако в целом на психологию все это подействовало опустошительно. Дело было не только в том, что разоблачались такие легковесные, эффектные исследования, как посвященные праймингу или позам силы, — огромное количество куда более “серьезных” работ по психологии (стэнфордский тюремный эксперимент и многие другие) тоже оказалось поставлено под сомнение. И проблема была не в том, что откопали какое-то ненужное старье и наглядно показали, что оно никуда не годится, — как когда папа Стефан VI в 897 году эксгумировал труп одного из своих предшественников, папы Формоза, и отдал под суд (тот был признан виновным). Нет, на работы, воспроизвести которые не удалось, продолжали как ни в чем не бывало ссылаться как ученые, так и писатели: целые направления исследований и пользующиеся хорошим

спросом научно-популярные книги строились на таком шатком фундаменте. Слово “кризис” кажется весьма точным описанием ситуации.

Мы можем попробовать утешиться тем, что в психологии как дисциплине есть нечто уникальное, что и вызвало ее кризис воспроизводимости. У психологов незавидная работа: они пытаются разобраться в крайне изменчивых и чрезвычайно сложных человеческих существах, со всеми их разными личностями, знаниями, опытом, настроениями и особенностями. Изучаемые психологами объекты, такие как мысли, эмоции, внимание, способности, восприятие, обычно неуловимы — их трудно, если вообще возможно, зафиксировать в лабораторном эксперименте. А в социальной психологии ученым приходится изучать, как все эти хитросплетенные люди друг с другом взаимодействуют. Не могла ли невероятная сложность задачи сделать открытия в психологии особенно ненадежными по сравнению с другими науками?

В этом аргументе есть кое-что справедливое: во многих исследованиях по психологии интересующее ученых явление едва затрагивается, тогда как другие, более “точные” науки, скажем физика, характеризуются лучше разработанными теориями и более точными и по-настоящему объективными измерениями. Однако нельзя сказать, что только в психологии есть проблемы с воспроизводимостью: хотя ни в какой другой области науки столь систематически и детально еще не изучалась доля успешно воспроизводящихся результатов, есть намеки на однотипные проблемы в огромном количестве разных направлений.

- Экономика: в исследовании 2016 года, повторяющем восемнадцать работ по микроэкономике (когда люди приходят в лабораторию и принимают участие в экспериментах, посвященных их экономическому поведению, — что не слиш-

ком отличается от исследований по психологии), доля воспроизводимости равнялась лишь 61 %[35].

- Нейронауки: в исследовании 2018 года обнаружилось, что стандартные работы по функциональной нейровизуализации, когда с помощью магнитно-резонансной томографии регистрируется активность мозга, пока человек выполняет какие-то задания (или просто лежит внутри МРТ-сканера), отличались лишь “незначительной воспроизводимостью”[36]. Еще мир функциональной нейровизуализации сотрясла статья, в которой вскрылось, что дефолтные настройки пакета программ, широко используемого для анализа данных визуализации, содержат статистическую ошибку. Это привело к громадному числу случайных нескорректированных ложноположительных результатов и скомпрометировало примерно 10 % всех статей, когда-либо опубликованных по этой теме[37].
- Эволюционная биология и экология: на целый ряд классических результатов, давно попавших в учебники и вызубриваемых поколениями студентов, посыпались критические обзоры после попыток их воспроизвести. Так, выяснилось, что заявления о знаменитом “синдроме одомашнивания”, когда лисицы в СССР, отбираемые по признаку дружелюбности, начинали приобретать внешний облик одомашненных видов (например, висячие уши и укороченные, широкие морды), были сильно преувеличены, причем большинство признаков “приручения” существовало еще до начала процесса селекции[38]. И многое из того, что, как мы думали, нам известно о половом отборе у птиц, было развенчано при получении более надежных данных. Скажем, в противоположность тому, что мы якобы знали, красная повязка на лапках у самцов зебровых амадин, похоже, *не* делает их сверхпривлекательными для самок; самцы воробьев с более крупным пятном черных перьев на горле (так называемым нагрудником), похоже, *не* доминируют в стае; а доказательства, что самок обыкновенных лазоревок больше привлекают определенные цвета оперения у самцов, неубедительны[39].

- Биология моря: в масштабном исследовании 2020 года, повторяющем другие работы, выяснилось, что закисление океана (как и изменение климата, это одно из последствий повышения уровня диоксида углерода в атмосфере) не влияет на поведение рыб[40]. Таким образом, не удалось воспроизвести несколько исследований предыдущего десятилетия, получивших широкую огласку, которые явно показывали, что в закисленной среде рыбы становятся дезориентированными и иногда плывут *по направлению* к химическим сигналам хищников, а не от них.
- Органическая химия: журнал *Organic Syntheses*, придерживающийся необычной политики — член редколлегии пробует повторить в собственной лаборатории результаты каждой подаваемой на рассмотрение статьи, — сообщил, что отказывает авторам 7,5 % работ из-за провалившихся попыток воспроизвести исследование[41].

Есть бесчисленное множество и других примеров: почти каждый случай, что я буду описывать в этой книге, содержит научное “открытие”, при ближайшем рассмотрении оказавшееся либо менее надежным, чем казалось, либо и вовсе недостоверным. Однако еще тревожнее то, что эти примеры порождены исследованиями, которые подверглись столь тщательному изучению, — получается, это лишь те примеры, о которых мы знаем. Сколько еще результатов, должны задаться мы вопросом, окажутся невоспроизводимыми, если кому-то случится попробовать их повторить?

Одна из причин, почему мы живем в такой неопределенности, заключается в том, что, как говорилось в предисловии, почти никто не проводит исследований, повторяющих прежние работы. Хотя в нашем распоряжении для большинства областей нет количественных данных, анализ специализированной литературы для некоторых из них позволяет сделать мрачный вывод. В экономике жалкие 0,1 % всех опубликованных статей посвящены попыткам воспроизвести предыдущие ис-

следования; в психологии этот показатель выше, но все равно весьма прискорбный — чуть больше 1 %[42]. Если все неустанно рвутся вперед к новым открытиям, не делая остановок, чтобы проверить, надежны ли уже имеющиеся знания, так ли уж удивителен приведенный выше список провалившихся попыток что-то воспроизвести?

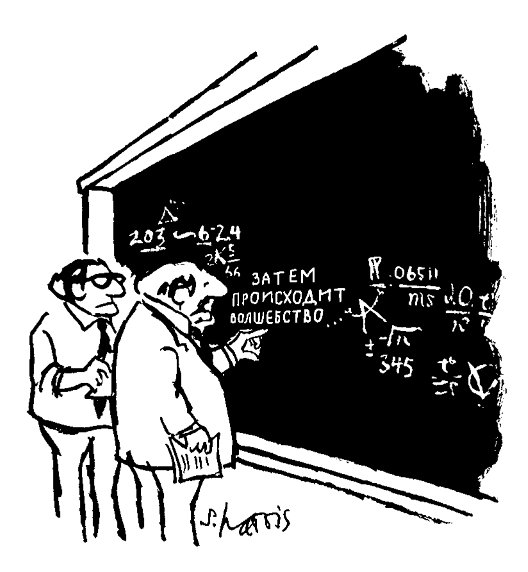
А вот что вызывает, пожалуй, еще большую озабоченность. Казалось бы, если вы получили точно такой же набор данных, как и в опубликованной ранее статье, вы сможете прийти к абсолютно тем же результатам, что описаны авторами. К сожалению, во многих областях исследователи сталкивались с невероятными трудностями при выполнении этой вроде бы нехитрой задачи. Иногда именно подобную проблему называют проблемой *воспроизводимости*, в противоположность проблеме *сходимости* результатов (последний термин обычно используется применительно к исследованиям, в которых ученые задаются теми же вопросами, но работают с другими данными)*. Как это возможно, чтобы результаты в таких условиях не воспроизвелись? Иногда причина в ошибках исходного исследования. А бывает и так, что авторы исходной работы недостаточно четко описали свой анализ, например, прибегали ко всяким выкрутасам со статистикой, о которых в статье не доложили, и поэтому их конкретные шаги независимые исследователи воссоздать не могут. Когда другие ученые как-то по-своему проводят статистический анализ данных, результаты выходят иные. Такие статьи — словно кулинарная книга, где полно фотографий блюд, от ко-

* Во избежание путаницы заметим, что в англоязычной специализированной литературе существуют два разных термина: “проблема *сходимости* результатов” *(replicability* или *repeatability)* и “проблема *воспроизводимости* результатов” *(reproducibility)*. В русскоязычной же литературе обычно не делается различий между этими случаями и используется единый термин — “проблема *воспроизводимости*”. Учитывая, что и в английском языке применение двух разных терминов не строгое (на что, в частности, указывает и сам автор в примечании 49 к этой главе), в русском переводе данной книги используется только термин “воспроизводимость”, тем более что необходимые детали соответствующих исследований там, где они важны, поясняются автором отдельно. *(Здесь и далее — прим. перев.)*

торых просто слюнки текут, но мало внимания уделено описанию ингредиентов и рецептам, необходимым для создания этих шедевров.

В макроэкономике (изучающей, например, налоговую политику и ее влияние на экономическое развитие стран) при повторном анализе шестидесяти семи статей ученые, используя точно такие же наборы данных, сумели воспроизвести результаты лишь двадцати двух, и последующее привлечение к работе авторов тех статей помогло несильно[43]. В науках о Земле исследователи испытывали как минимум небольшие трудности при получении тех же результатов в случае тридцати семи из тридцати девяти изучавшихся ими статей[44]. А когда исследователи машинного обучения проанализировали набор статей об "алгоритмах рекомендаций" (это тип компьютерных программ, которые используются сайтами вроде *Amazon* и *Netflix*, чтобы на основании того, что люди вроде вас выбирали раньше, предугадывать, какую покупку вам сейчас захотелось бы сделать или какой фильм посмотреть), то смогли воспроизвести только семь из восемнадцати работ на эту тему, незадолго до того представленных на престижных конференциях по компьютерным системам[45]. Те статьи — воплощение классической карикатуры Сидни Харриса.

Вы вправе удивиться, почему некоторые из перечисленных выше примеров вообще имеют значение. Хоть мы и наблюдали плохую воспроизводимость в кое-каких важных областях, например в экономической теории, каким образом наша жизнь может измениться, если кучка ученых в итоге разойдется во взглядах на то, работают ли позы силы и отличаются ли альфа-самцы воробьев более крупным пятном черных перьев? На этот вопрос есть два ответа. Первый заключается в том, что на чашу весов положен более общий принцип: наука критически важна для нашего общества, и мы не должны допускать появления низкокачественных, невоспроизводимых исследований, компрометирующих ее, ни в одной области. Если мы позволим стандартам в любой области просесть,



“Думаю, вам следует подробнее расписать второй шаг”.

мы рискуем испортить репутацию науки в целом. Второй ответ связан с научным направлением, которое мы еще не рассматривали, где отсутствие воспроизводимости имеет бесспорные прямые последствия. Это, конечно же, область медицинских исследований.

Примерно в то время, когда кризис воспроизводимости назревал в психологии, ученые из биотехнологической компании *Amgen* попробовали повторить пятьдесят три ключевых “доклинических” исследования рака, результаты которых были опубликованы в топовых научных журналах (доклинические исследования — это те, что проводятся на первых этапах разработки лекарства, обычно на мышах или на человеческих клетках *in vitro**)[46]. Всего шесть из этих попыток (а это лишь 11 %) увенчались успехом. Итог сходных попыток другой фирмы, *Bayer*, оказался немногим лучше — около 20 %[47]. Подобное отсутствие строгого подкрепления результатов

* То есть в пробирке.

в области доклинических исследований — вероятно, одна из причин, по которым испытания лекарств от рака так часто разочаровывают: согласно одной оценке, только 3,4 % таких лекарств проходят весь путь от первых доклинических исследований до применения на людях[48].

Подобные неприятные открытия заставили ученых, занимающихся раком, как и психологов, задуматься об общем состоянии их области. В 2013 году они объединились для совместной попытки повторить пятьдесят важных доклинических исследований рака в независимых лабораториях[49]. В тех работах утверждалось, например, что конкретный вид бактерий может быть связан с ростом опухоли при колоректальном раке и что некоторые мутации при лейкемии связаны с активностью определенного фермента[50]. Но еще до начала проведения исследований-повторений возникли трудности. В каждой без исключений исходной статье для всех до единого экспериментов авторы приводили недостаточно сведений для того, чтобы читатель мог понять, как провести точно такой же опыт[51]. Технические детали экспериментов — например, плотность используемых клеток или особенности измерений и анализа — попросту не упоминались. Работа по повторению исследований забуксовала, поскольку потребовала обширной переписки с авторами исходных статей, которым для отыскания конкретных деталей своих экспериментов зачастую приходилось откапывать старые лабораторные журналы и связываться с бывшими коллегами[52]. Кто-то сотрудничать не хотел: о 45 % авторов исходных статей исследователи отозвались как о “минимально полезных” или “вообще не оказавших помощи”[53]. Возможно, они беспокоились, что ученые, повторяющие их работу, окажутся некомпетентными или что их будущие исследования не получат финансирования, если воспроизвести их прежние результаты не удастся[54].

Позже в более масштабном исследовании были случайным образом выбраны двести шестьдесят восемь биомедицин-

ских статей, в том числе и из области клинических испытаний. Обнаружилось, что ни в одной из них, за *единственным* исключением, не содержался полный протокол исследования. А это означало опять-таки, что даже для того, чтобы предпринять *попытку* повторить эксперименты, требовалось знать дополнительные детали, не описанные в статье[55]. Другой анализ выявил, что в 54 % биомедицинских статей даже не описывалось толком, какие животные, реагенты или клетки использовались в экспериментах[56]. Давайте на минуту задумаемся, насколько же это странно. Если в статье исследование описывается лишь в общих чертах, а необходимые детали приходится месяцами выуживать из электронной переписки с авторами (а то они и вовсе оказываются навеки утеряны), зачем вообще было ее писать? Вернувшись мысленно хотя бы в XVII век к Роберту Бойлю, вспомним, что изначальная, глубинная цель ученых состояла в том, чтобы докладывать о каждой мелочи в своих штудиях, так чтобы другие могли досконально все изучить и попробовать повторить их исследования. Упомянутые статьи провалили эту фундаментальную проверку, равно как и опубликовавшие их журналы не сумели выполнить свою базовую, важнейшую функцию.

В проекте по воспроизведению исследований рака все эти проблемы с попытками повторить эксперименты вкупе с некоторыми финансовыми затруднениями привели к тому, что ученым пришлось постепенно сократить число работ, которые они намеревались повторить, — с пятидесяти до всего лишь восемнадцати[57]. К моменту написания этих строк отчет о повторении результатов четырнадцати из них уже опубликован, и картина получилась смешанная: для пяти исходных статей важные результаты (включая связь между лейкемией и ферментом) четко воспроизвелись, для четырех — воспроизвелись частично, для трех — совсем не воспроизвелись (в том числе связь между бактериями и колоректальным раком), а для двух результаты даже не получилось интерпретировать[58]. Повторять, стоит отметить, непросто.

Проблемы с воспроизводимостью в медицине затронули не только лабораторные доклинические исследования — они могут напрямую влиять на средства лечения, прописываемые врачами своим пациентам. Оказывается, широко распространенные средства лечения часто основаны на низкокачественных исследованиях: вместо того чтобы прочно уходить корнями в доказательства, общепризнанная медицинская мудрость регулярно вступает в противоречие с результатами новых исследований. Подобное происходит так часто, что ученые-медики Винай Прасад и Адам Сифу окрестили это явление “медицинской реверсией”[59].

Один особенно поразительный пример медицинской реверсии связан с “интранаркозным пробуждением”. Это неброское название дали жуткому (но, благо, редкому) осложнению, когда человек пробуждается во время операции, иногда чувствуя невыносимую боль от рассечения собственных тканей, и не может ни двигаться, ни говорить, ни сделать что-либо еще. Исследования 1990-х годов обосновали использование прибора под названием “монитор биспектрального индекса” (или “BIS-монитор” — от английского словосочетания *bispectral index*). По сути это электрод, который крепится к коже головы и позволяет хирургам удостовериться, что пациент действительно без сознания. Исследования воплотились в широко распространенную практику: к 2007 году половина операционных в США обзавелась таким прибором, и с ним было проведено около сорока миллионов операций по всему миру[60]. Однако выяснилось, что с исходными работами не все в порядке. Когда в 2008 году провели более масштабное и более высококачественное исследование, обнаружилось, что BIS-монитор бесполезен: “интранаркозное пробуждение случалось, даже когда [приборные] показания… укладывались в допустимый диапазон”[61].

В 2019 году Прасад, Сифу и их коллеги проанализировали более трех тысяч статей из трех самых престижных медицинских журналов и обнаружили, что не менее

чем в трехстах девяноста шести из них опровергались устоявшиеся в медицинской практике мнения[62]. Вот лишь несколько примеров:

- *Роды*. В некоторых предыдущих исследованиях было показано, что, когда женщина рожает двойню, плановое кесарево сечение — самый безопасный вариант для младенцев. В результате это стало общепринятой практикой (по крайней мере в Северной Америке). Но в крупном рандомизированном исследовании 2013 года ученые не выявили никакой разницы с точки зрения здоровья детей[63].
- *Аллергия*. Аллергия на арахис бывает смертельной, и если у родителя она есть, то высок риск, что она разовьется и у его детей. Долгие годы рекомендации для младенцев из группы риска, основанные на результатах прежних исследований, предписывали не давать детям арахис как минимум до трех лет и кормящим матерям тоже его избегать. Оказывается, этот совет был неверным, все как раз наоборот: добротные рандомизированные испытания 2015 года показали, что лишь у примерно 2 % детей из группы риска, евших арахис в первые годы жизни, к пяти годам развилась на него аллергия, тогда как среди детей из группы риска, не употреблявших арахис, аллергия появилась почти у 14 %[64].
- *Инфаркт миокарда*. В некоторых небольших исследованиях было показано, что шансы выжить при остановке сердца у человека повышаются, если на несколько градусов снизить его температуру тела. Рекомендацию, основанную на этом открытии, начали включать в руководства для парамедиков. Однако в крупном исследовании 2014 года выяснилось, что охлаждение не влияет на шансы выжить, а то и вообще, возможно, повышает вероятность второго инфаркта при транспортировке пациента в больницу[65].
- *Инсульт*. На основании исследований предполагалось, что человека, перенесшего инсульт, лучше всего как можно раньше заставить двигаться: садиться в кровати, стоять, хо-

> дить, если возможно. Концепция "ранней мобилизации" встречается во многих широко используемых больничных методичках. Однако в масштабном рандомизированном исследовании 2015 года было показано, что ранняя мобилизация на самом деле вела к более *неблагоприятным* исходам для пациентов с инсультом[66]. Аналогично в исследовании 2016 года выяснилось, что принятая практика переливать пациентам с инсультом тромбоциты (это процедура, восполняющая запасы клеточных элементов, участвующих в процессе свертывания крови, что в теории помогает предотвратить дальнейшее кровотечение) в действительности все только ухудшала[67].

Вполне понятно, почему врачи и авторы клинических рекомендаций иногда невольно полагаются на малодостоверные доказательства. Зачастую альтернатива — это вообще отсутствие доказательств, а ведь их долг — помочь пациентам, нуждающимся в лечении, *прямо сейчас*. А развитие технологий, методик и финансирования позволяет ученым сегодня проводить более качественные исследования, чем было возможно несколько лет назад, — научный прогресс так и происходит, ничего не поделаешь. Только вот ученые подвели врачей и пациентов, поскольку, проводя низкокачественные эксперименты и публикуя их результаты, создали такую неопределенность в медицинской литературе, что и студентам курсов по методологии исследований понятно: это ненормально. Даже на тот момент, когда эти исходные статьи выходили, мы знали, как сработать лучше, — и однако же этого не сделали.

Степень неопределенности в медицинской науке можно оценить в полной мере, если взглянуть на всю медицинскую литературу разом. Один из способов это сделать — проанализировать множество всеобъемлющих обзоров, опубликованных солиднейшей некоммерческой организацией "Кокрейн", которая систематически оценивает качество средств лечения. В пугающих 45 % таких обзоров авторы заключают: для того

чтобы решить, работает или нет обсуждаемое средство лечения, данных недостаточно[68].

Сколько пациентов напрасно обнадежили, обрекли на страдания или даже на смерть, потому что их врачи использовали бесполезные или губительные средства лечения, которые лишь *казались* опирающимися на науку? А если отвлечься от человеческих страданий — представьте, сколько денег было разбазарено! Если считать, что лишь половина доклинических испытаний воспроизводима — предположение обоснованное, хотя, конечно, спорное, — и вычислить сумму денег, растрачиваемую ежегодно на низкокачественные исследования, которые нельзя повторить, то получится двадцать восемь миллиардов долларов только в США (включая финансирование фармацевтических компаний, правительственные гранты и вложения из других источников)[69]. Другие оценки существенно выше[70]. Даже если доля воспроизводимых исследований куда больше 50 %, все равно баснословные деньги выбрасываются на плохие исследования. Хуже того, эти вычисления учитывают только доклинические испытания, а еще больше денег будет выкинуто на дальнейших этапах, в последующих исследованиях — таких как испытания лекарств на людях, — которые выстраиваются на столь ненадежных основаниях. И приведенный выше расчет включал только стоимость самого исследования — не учитывались дополнительные потери, связанные с выводом на рынок для использования миллионами пациентов неэффективного средства лечения, например BIS-монитора для случаев интранаркозного пробуждения.

При всех этих провалах и “реверсиях”, вместе взятых, неудивительно, что столько ученых так беспокоятся об уровне воспроизводимости в своей области. В опросе с участием свыше полутора тысяч исследователей, проведенном в 2016 году (хоть, пожалуй, и не репрезентативном по-настоящему, поскольку

он учитывал лишь тех, кто заполнил опросник на сайте журнала *Nature*), выяснилось, что 52 % респондентов считают: разразился “серьезный кризис” воспроизводимости. А 38 % опрошенных думают, что царит по меньшей мере “легкий кризис”, как бы странно это ни звучало[71]. Около 90 % химиков сообщили, что им случалось не суметь повторить результат какого-то другого исследователя, примерно 80 % биологов сообщили о том же, как и почти 70 % физиков, инженеров и ученых-медиков. Лишь чуть меньше ученых в процентном соотношении сообщили, что сталкивались с проблемами при повторении своих *собственных* результатов. Поскольку это не было социологическим исследованием мнения ученых в полном смысле слова и те, кто уже тревожился по поводу воспроизводимости, вероятно, чаще проходили опрос, приведенные оценки представляются несколько преувеличенными. Однако они ясно говорят об одном: существуют широко распространенные опасения насчет того, какой части научной литературы, включая даже работы, выполненные нами собственноручно, можно доверять.

Мы должны были это предвидеть. Вышедшая в 2005 году статья Джона Иоаннидиса, занимающегося метанаукой, носила драматическое название “Почему большинство опубликованных результатов исследований неверны”, и представленная там математическая модель выявила именно это: как только вы принимаете в рассмотрение, сколько есть способов, какими научные исследования могут пойти не так, любое конкретное утверждение в научной статье с большей вероятностью оказывается ложным, а не правильным[72]. Хоть статья и привлекла много внимания, породив немало споров (на нее сослались более восьмисот раз в первые пять лет после публикации), но с точки зрения каких-то необходимых мер, которые бы ученые начали предпринимать для повышения качества исследований, то были предсказания в духе Кассандры, к коим все остались глухи[73]. Только после неприятных открытий, затрубивших о кризисе воспроизводимости, —

а начались они с обнародования Бемом парапсихологических заявлений и с разоблачения стапеловского мошенничества в 2011 году (и примерно в то же время провалились попытки повторить работы по праймингу и исследования рака) — проблема стала общепризнанной и пришло беспощадное осознание, что гнездится она в самой сути того, как сегодня делается наука[74].

Так как же мы достигли черты, за которой провокационное название вроде “Почему большинство опубликованных результатов исследований неверны” кажется не каким-то абсурдным преувеличением, а обоснованным предположением? Разберем теперь, какими многочисленными способами наука может пойти — и идет — по ложному пути.

ЧАСТЬ II
Ошибки и изъяны

Глава 3
Мошенничество

> Пусть мы не жаждем обнаруживать мошенничество, но мы не должны по этой причине пытаться не замечать его… Люди страшатся подлецов, но еще — не выносят глупцов.
>
> Норман Макдональд
> *“Максимы и размышления о морали” (1827)* [1]

Одни из самых поистине трогательных видео в интернете — это записи, на которых жизнь людей с болезнями или ограниченными возможностями вмиг меняется благодаря какому-то новому технологическому продукту. Малыши благодаря кохлеарному импланту, к своему удивлению и радости, обнаруживают, что впервые слышат звуки. Детям, родившимся с катарактой обоих глаз, проводят операцию, необходимую, чтобы они начали видеть. Военные, потерявшие в бою ноги, делают первые шаги с новыми протезами[2]. Эти видео распространяются по социальным сетям как пожар, и не зря: они не только трогательные, но еще и служат напоминанием о силе науки, способной — в идеале — улучшить наше здоровье и нашу жизнь.

Но вот история о том, как даже настолько чистая цель науки может быть извращена и изуродована: пациенты, думавшие, что их лечат по новейшей, передовой методике, в итоге оказались жертвами одного из самых безнравственных мошенников от науки, каких мы только видели в этом столетии. Хуже того, это была не афера какого-то шарлатана от “альтернативной” медицины, выцепляющего отчаявшихся пациентов в интернете. Это было мошенничество, совершенное в стенах самых уважаемых в мире медицинских институтов и на страницах престижнейших научных журналов. Мы увидим, что

даже самые бессовестные мошенники иногда могут прятаться у всех на виду.

Когда трахея (дыхательное горло) из-за болезни или травмы серьезно повреждается, хирурги не в силах снова соединить вместе ее обломки, и тогда единственный способ спасти пациента — дать ему новую трахею[3]. Трансплантировать трахею, как и любой большой орган, чрезвычайно трудно: сложность не только в том, чтобы найти потенциальных доноров (которые по понятным причинам должны быть мертвы), но и в том, что трахея обычно отторгается иммунной системой реципиента, если донор от него генетически отличен. Поэтому, вместо того чтобы использовать трансплантаты, хирурги десятилетиями пытались имплантировать своим пациентам искусственные трахеи, сделанные из потрясающе разнообразных материалов: пластика, нержавеющей стали, коллагена, даже стекла. Однако эти попытки почти всегда проваливались: синтетическое дыхательное горло смещалось, закупоривалось и инфицировалось. К началу XXI века в медицине сложился консенсус: вариант с искусственной трахеей нежизнеспособен[4].

И тут появляется Паоло Маккиарини, итальянский хирург, в 2008 году опубликовавший в самом престижном медицинском журнале *The Lancet* ставшую суперизвестной статью о своей успешной трансплантации трахеи[5]. Новшество Маккиарини состояло в том, чтобы до трансплантации “засеять” донорскую трахею стволовыми клетками (которые могут делиться бесконечно, исправляя повреждения и заменяя другие клетки организма) из образца, взятого у реципиента. После того как стволовые клетки некоторое время держали в специально сконструированном инкубаторе, они “колонизировали” донорскую трахею и таким образом вроде бы предотвращали ее отторжение во время последующей трансплантации. То был большой шаг вперед. Но святым Граалем оставалось создание полностью искусственной трахеи, вообще не требующей наличия донора. Могла ли идея Маккиарини — заставить тело

лучше принимать инородные предметы, вырастив вокруг них внешний слой совместимых с организмом клеток, — помочь наконец достигнуть этой цели?

Всего через несколько лет оказалось, что ответ — “да”. С момента выхода той его статьи репутация Маккиарини как хирурга-гения окрепла, и в 2010 году его пригласили на работу — по рекомендации четырнадцати уже работавших там профессоров — Каролинский институт Швеции, где Маккиарини стал приглашенным профессором, и больница при этом институте, где его сделали ведущим хирургом. Каролинский институт — не просто лидирующий в стране, где много отличных университетов, он присуждает Нобелевскую премию по физиологии и медицине. Столь авторитетному учреждению логично было нанять Маккиарини — хирурга, производящего революцию в регенеративной медицине своими искусными методами работы со стволовыми клетками.

В июле 2011 года Каролинский институт с волнением объявил, что сделан следующий шаг: в больнице при институте Маккиарини только что “впервые в истории” успешно трансплантировал пациенту с онкологией полностью синтетическую трахею из углерода и кремния, засеянную стволовыми клетками[6]. В ноябре того же года вышла научная статья с описанием подробностей операции, и к тому времени Маккиарини провел подобную трансплантацию еще одному пациенту больницы[7]. Статья, тоже опубликованная в журнале *The Lancet*, описывала “веские доказательства” успешности трансплантации. За 2012 год Маккиарини провел операции с искусственными трахеями еще на трех пациентах, одну на базе Каролинского института и две — во втором его операционном центре в Краснодаре в России. Еще две трансплантации в России должны были состояться в течение ближайших двух лет, и Маккиарини распространял хорошие новости, публикуя всё новые научные статьи[8].

В одной из них, вышедшей в 2014 году в журнале *Biomaterials* и полной симпатичных фотографий “клеточного каркаса

трахеи, полученного методом электроформования" с электронного микроскопа, авторы кратко признавали, что у первого пациента возникли некоторые проблемы, и тут же продолжали живописать чудо новой технологии. Только вот они упустили одну обескураживающую деталь: пациент умер — за семь недель до того, как статью вообще приняли к публикации[9]. Пациент, прооперированный вторым, умер еще раньше, спустя всего лишь три месяца после трансплантации[10]. Третий пациент из больницы при Каролинском институте умрет в 2017 году, после серии неудачных последующих операций[11]. Российским пациентам повезло не больше. Первая пациентка, Юлия Туулик, артистка балета из Санкт-Петербурга, описала журналисту свое ужасное состояние:

> Все у меня очень, очень плохо. Я пролежала больше шести месяцев в краснодарской больнице. Прошла через тридцать с лишним операций под общим наркозом. Через три недели после первой операции открылся гнойный свищ [отверстие, из которого сочится гной], и с тех пор моя шея гниет. Я вешу сорок семь килограммов. Едва могу ходить. Мне трудно дышать, и сейчас у меня нет голоса. И воняет так… что люди отшатываются[12].

Туулик умерла в 2014-м, через два года после операции[13]. Самое душераздирающее — ее состояние до операции даже не было угрожающим жизни[14]. Второй российский пациент погиб в "велосипедной аварии", как это было описано, еще один умер при невыясненных обстоятельствах через год после операции, а другой выжил, но только благодаря тому, что синтетическую трахею удалили[15]. Еще Маккиарини провел трансплантацию маленькому ребенку канадско-южнокорейского происхождения в больнице США в городе Пеория, штат Иллинойс, в 2013 году под пристальным вниманием средств массовой информации. Малышка умерла всего несколько месяцев спустя[16].

Группа врачей из больницы при Каролинском институте, наблюдавшая пациентов Маккиарини после операций, не смогла увязать между собой их ужасающее состояние и хвалебное описание результатов в научных статьях. Они посовещались и подали совместную жалобу руководству Каролинского института. Вместо изумления и беспокойства они столкнулись с полным игнорированием, их пытались заставить замолчать. Институт даже сообщил о врачах в полицию: мол, просматривая медицинские записи пациентов, те нарушили политику конфиденциальности (эти обвинения были быстро сняты)[17]. Однако в конце концов институтская верхушка поддалась давлению и пригласила независимого исследователя — профессора из расположенного неподалеку Уппсальского университета — расследовать эти заявления.

Его отчет, подготовленный к маю 2015 года и насчитывавший двадцать тысяч слов, показал абсолютно однозначно: Маккиарини "виновен в недобросовестной научной практике" по нескольким пунктам[18]. В семи своих статьях Маккиарини неправомерно утверждал, что состояние пациентов улучшалось, хотя необходимые проверки даже не проводились; неверно указывал периоды наблюдения, чтобы казалось, будто здоровье пациентов дольше было хорошим; не сообщал, что пациенты страдали от тяжелых осложнений и иногда вынуждены были ложиться на последующие операции; не добыл нужного разрешения от этического комитета на проведение, по сути, медицинских экспериментов на людях; фальсифицировал данные лабораторного исследования, где он заменял пищеводы крысам на искусственные[19].

Можно было бы подумать, что история на том и закончится. Однако после того, как Маккиарини ответил на обвинения, изложенные в отчете о независимом исследовании, Каролинский институт решил провести собственное, внутреннее расследование. В августе 2015 года на основании информации, которая не была обнародована, институт пришел к заключению, что на самом-то деле никаких нарушений на-

учной этики не было[20]. На следующей неделе в журнале *The Lancet* вышла редакционная заметка, восславляющая тот факт, что “Паоло Маккиарини не виновен в недобросовестной научной практике”[21]. Так с Маккиарини самым очевидным образом были сняты все обвинения — причем двумя мощнейшими столпами медицинского мира.

Затем, в январе 2016 года, всплыли два поразительных обстоятельства, которые нельзя было проигнорировать или замять. Во-первых, в журнале *Vanity Fair* появилась статья, подробно описывающая роман Маккиарини с продюсером *NBC News* Бенитой Александер, в 2014-м он делал ей предложение[22]. Он заверял ее, что является личным врачом папы Франциска и пригласит понтифика к ним на свадьбу, как и некоторых ярчайших знаменитостей, включая чету Обама, Рассела Кроу и Элтона Джона[23]. Но когда сотрудники *Vanity Fair* связались с Ватиканом, им сообщили, что у папы нет и не было никакого врача по фамилии Маккиарини. Мало того, выяснилось, что все время своих отношений с Александер Маккиарини был женат на ком-то еще — и у него было двое маленьких детей[24].

Вторым обстоятельством стал выход шведского документального фильма в трех частях под названием “Эксперимент”, в котором до ужаса детально показано, как жизнь пациентов Маккиарини была разрушена, а иногда оборвана из-за его вопиющей некомпетентности. В фильме есть фрагмент с бронхоскопией (когда крошечную камеру спускают вниз по трахее), которую делали первому пациенту из больницы при Каролинском институте. В ходе процедуры обнаружилось, что трахея вся в рубцах, закупорена и даже со сквозными отверстиями — совсем не те “почти нормальные дыхательные пути”, которые описывал Маккиарини в своей статье для журнала *The Lancet*[25].

Это хорошенько встряхнуло Каролинский институт, и он развернул совершенно новое расследование. На сей раз полетели головы. Ректор, с самого начала поддерживавший Маккиарини, подал в отставку. Так же поступил декан по науч-

ной работе, затем председатель совета университета, следом член Нобелевского комитета, который пробивал назначение Маккиарини[26]. В марте 2016 года Маккиарини был наконец отстранен от должности[27]. С первой преступной операции с синтетической трахеей к тому времени прошло семь лет.

После своей упорной обороны всего несколькими годами ранее журнал *The Lancet* отозвал статьи Маккиарини о синтетических трахеях[28]. Институт в середине 2018 года подтвердил, что обнаружил нарушения научной этики в еще нескольких статьях Маккиарини, список которых вывесил в интернете; все они были отозваны[29]. После увольнения Маккиарини ретировался в Россию и продолжил свои "исследования"[30]. К счастью, в его самой последней опубликованной статье речь о пациентах не идет, вместо этого он и его новые коллеги проверяли совместимость пластикового пищевода с клетками, взятыми у мертвых павианов[31]. Текущий статус Маккиарини неясен: российское правительство отозвало финансирование в середине 2017 года, так что вряд ли он умудрится провести еще хоть одну операцию[32]. И возможно, все-таки восторжествует некая справедливость для его пациентов: в декабре 2018 года шведская прокуратура сделала заявление о том, что возобновляет расследование против Маккиарини по двум случаям непредумышленного убийства*[33].

В чем причина столь затянувшегося процесса, когда даже ужасная, мучительная смерть нескольких пациентов не сподвигла институт уволить или хотя бы подвергнуть порицанию ответственного за все это человека? И вообще, почему институт обрушился с нападками на тех самых людей, которые забили тревогу из-за фальсификации результатов?[34] Определенно, Маккиарини использовал свои "прорывные" операции, чтобы раздуть собственную репутацию и славу. Но еще Маккиарини был козырем Каролинского института с его пла-

* Интересующихся последними новостями о Маккиарини отсылаем на сайт Каролинского института, где представлена хронология событий: *https://news.ki.se/the-macchiarini-case-timeline.*

нами на расширение за рубежом. Предполагалось, что связь со светилом хирургии поможет обеспечить беспрепятственное создание нового центра регенеративной медицины, который институт планировал открыть в Гонконге[35]. Полнейшая паника и стыд руководства института тоже, вероятно, послужили причиной желания замять дело: такое солидное учреждение едва ли жаждет признаваться самому себе, не говоря уже об общественности, что подпустило смертельно опасного мошенника к уязвимым пациентам.

Немного найдется случаев мошенничества в науке, воздействовавших на человеческие жизни столь ужасающе прямо, как искусственные трахеи. Мало кто из мошенников от науки столь вопиюще бессовестен, как Маккиарини. И все же из его истории мы можем извлечь несколько общих уроков. Во-первых, мы видим, как сильно наука, несмотря на заложенный в нее организованный скептицизм, сводится к доверию: что исследования на самом деле проводились так, как о том сообщается, что числа действительно те, которые выдал статистический анализ, и, как в обсужденном выше случае, что пациенты и правда выздоравливали, как о том заявлено. Мошенничество демонстрирует, насколько злокозненно этим доверием можно спекулировать. Во-вторых, очевидно, что с такой же меркой скептицизма, с какой мы подходим к исследованиям и людям, нужно подходить еще и к учреждениям. Всегда будут негодяи, чья жажда славы и успеха перевешивает все остальное, но нам нужно знать, что известные научные учреждения вроде Каролинского института и журнала *The Lancet* делают все возможное, чтобы не допускать влияния этих подлецов на науку — и обличать и наказывать их, когда бы они ни подняли голову. Увы, желание учреждений престижа ради нанимать ярких ученых и публиковать их работы может вылиться в то, что институты и журналы станут добровольно закрывать глаза на действия мошенников, а иногда будет доходить до того, что они станут даже защищать негодяев от последствий их действий.

Или еще хуже: то обстоятельство, что научное сообщество так гордится и дорожит своим образом объективной и безупречно честной системы — где мошенничество грозит анафемой, — может, как это ни противоестественно, служить препятствием для вычисления злоумышленников в собственных рядах. Сама мысль о том, что в науке существуют такие негодяи, как Маккиарини, столь мучительна, что многие занимают позицию “не вижу зла”, не замечая даже самых явных, бросающихся в глаза признаков нарушения научной этики. А кто-то ударяется в отрицание распространенности мошенничества и наличия у него последствий. Однако, как мы увидим в этой главе, мошенничество в науке — не исчезающе редкий сценарий, как бы отчаянно нам того ни хотелось. На самом деле оно происходит удручающе часто.

Один из самых известных (и абсурднейших) случаев научного мошенничества XX века тоже связан с трансплантатами — кожными лоскутами. Дерматолог Уильям Саммерлин, в 1974 году работавший в престижном Институте исследований рака имени Слоуна — Кеттеринга в Нью-Йорке, предвосхитил идею Маккиарини, заявив, что разрешил проблему отторжения трансплантатов, которую мы обсуждали выше. Применив обезоруживающе простую новую методику, когда кожа донора перед операцией инкубируется, промариновываясь в специальной питательной среде, Саммерлин будто бы пересадил лоскут кожи от черной мыши белой — без всякого иммунного отторжения. Только вот он этого не сделал. Перед тем как показать заведующему лаборатории свои впечатляющие новые результаты, он подкрасил кусочек меха белой мыши черным маркером — эту уловку позже обнаружил лаборант, который, почуяв неладное (запах мыши в данном случае), стал спиртом оттирать с грызуна краску. Ни единого раза пересадка кожи на мышах не увенчалась успехом, и Саммерлина вскоре уволили[36].

Едва ли Саммерлин единственный среди ученых, кто потворствовал своим недозволенным артистическим порывам. Как правило, это видно по рисункам, иллюстрирующим научные статьи. Благодаря компьютерной графике чего уж проще — обрезать, дублировать, подрисовывать, склеивать, перекрашивать или как-то еще изменять изображения, полученные в ходе научных исследований, так чтобы они показывали все, что вам хочется. Разумеется, производство жульнических фотографий было более чем возможно задолго до наступления эры фотошопа (вспомним хотя бы знаменитое "исчезновение" комиссара Николая Ежова с фотографии, где он был запечатлен рядом с Иосифом Сталиным, после того как впал в немилость у советского лидера). В 1961 году журнал *Science* принес извинения за публикацию статьи индийских ветеринаров, утверждавших, будто они впервые обнаружили паразита *Toxoplasma gondii* в куриных яйцах (потенциальный риск для здоровья человека, поскольку этот паразит вызывает токсоплазмоз — опасное заболевание для тех, у кого ослаблена иммунная система)[37]. Доказательство присутствия паразита — фотографии его цист в курином яйце, полученные с помощью микроскопа, — было, как выяснилось, поддельным. Две разные, по словам исследователей, цисты оказались на деле одним и тем же изображением, которое уменьшили и отразили по горизонтали — невыносимо банальное дублирование, ясно видное теперь, но тогда не замеченное рецензентами. Вскоре после того, как этот обман раскрылся, исследователей или заставили уволиться, или отстранили от занимаемых должностей[38].

Возможно, вы думаете, что только самые ленивые мошенники от науки прибегают в своих статьях к дублированию изображений, делая свои злоухищрения видимыми для бдительного невооруженного глаза. Однако дублирование изображений всплывает снова и снова, этот прием стал основной особенностью ряда крупнейших случаев мошенничества последних десятилетий. В 2004 году биолог Хван У Сок из Юж-

ной Кореи на страницах журнала *Science* объявил о том, что успешно клонировал человеческие эмбрионы. В следующем году в том же журнале он заявил, что получил — из тех эмбрионов — первые клонированные линии стволовых клеток. Стволовые клетки столь многообещающи потому (помимо того, что могут продолжать делиться бесконечно), что они "плюрипотентны", то есть умеют трансформироваться во многие разные типы тканей (нейроны, клетки печени, клетки крови и так далее) — в лаборатории они словно швейцарский армейский нож. С помощью клонированных клеточных линий — в статье Хвана сообщалось, что их создано одиннадцать, — стало бы возможно производить персонализированные средства лечения на основе стволовых клеток, а значит, восстанавливать поврежденные ткани и травмированные или пораженные болезнью органы. Как и в случае с трансплантатами трахеи, это означало бы, что персонализированные стволовые клетки получены от самого пациента и что его иммунная система с меньшей вероятностью отторгнет какие-либо средства лечения на основе таких клеток. В еще одной прорывной статье того же года группа Хвана из Сеульского национального университета явила миру первую клонированную собаку — афганскую борзую по кличке Снаппи[39].

Трудно себе представить, насколько знаменитым стал Хван в Южной Корее благодаря этим достижениям. В средствах массовой информации перед ним просто благоговели[40]. На улицах и на общественном транспорте появились постеры с изображением его лица и надписями вроде "Надежда мира — мечта Кореи"[41]. Почта в 2005 году выпустила специальную марку, прославляющую его работу: на ней (пожалуй, весьма преждевременно) человеческий силуэт вставал с кресла-каталки, подпрыгивал и обнимал любимого[42]. Корейское правительство, прозвавшее Хвана "величайшим ученым", вбухало в его исследования громадные суммы. Сотни женщин устремились к Хвану, чтобы пожертвовать свои яйцеклетки на его исследования[43].

Угадайте, что было дальше. По мере изучения статьи в *Science* выяснилось, что два изображения, якобы показывающие клеточные линии Хвана для разных пациентов, идентичны (поясним: шансы, что два изображения подобного рода окажутся идентичными по чистой случайности, фактически нулевые). Еще всплыло задвоение на двух других рисунках: фрагменты одной и той же фотографии подавались как совершенно разные[44]. Будь это все проблемы, их легко можно было бы списать на недосмотр — просто кто-то перепутал или неверно подписал фотографии. Но какое там! Разоблачители из лаборатории Хвана рассказали, что созданы были лишь две клеточные линии, а не одиннадцать и ни одна из них не имела отношения к клонированным эмбрионам[45]. Остальные фотографии клеток были подделаны или специально названы неправильно по указаниям Хвана. Весь исследовательский проект был пшиком.

Еще до того, как махинации с изображениями вызвали обеспокоенность, Хван попал в затруднительное положение, принимая яйцеклетки от доноров, которые не были должным образом проинформированы, как их биоматериал станут использовать и какие потенциальные риски они несут из-за изъятия яйцеклеток. Хван также заставлял сотрудниц своей лаборатории жертвовать их собственные яйцеклетки на эксперименты[46]. И обнаруживались все новые свидетельства недобросовестной научной практики: Хван перевел часть финансирования, предназначавшегося для его исследований, на контролируемые им банковские счета — и хотя он утверждал, что деньги все равно потрачены на научный прибор, в ходе расследования выяснилось, что “прибор” этот включал в себя новую машину для жены Хвана и отчисления политикам, оказывающим ему поддержку[47]. Средства массовой информации в целом с таким восторгом следили за Хваном, что даже это грубое научное мошенничество не охладило пыл его почитателей. Протестующие толпились на улицах перед офисами медиакомпаний, которые выпускали какие-то негативные

материалы, и бомбардировали их онлайн-форумы тысячами постов гневного содержания в поддержку Хвана[48]. В итоге властям пришлось вмешаться. Хвана уволили из университета, а затем он подвергся уголовному преследованию, хотя и ухитрился избежать тюрьмы, получив два года условно[49]. Сейчас он все еще занимается клонированием, но в менее престижном университете, и его работа привлекает лишь толику былого внимания. Кстати, среди всех фальшивых хвановских достижений Снаппи был настоящим: анализ ДНК показал, что он действительно клон другой афганской борзой по кличке Тай. Снаппи умер в 2015 году, но в каком-то смысле он все еще с нами: четыре уже его клона родились двумя годами позже[50].

Несомненно, Хван был знаменитейшим ученым в своей стране и одним из самых выдающихся биологов в мире. При всем том внимании, которое ему уделялось, почему он думал, что ему сойдет с рук столь вопиющий, неприкрытый обман? Ответ на этот вопрос имеет отношение не только к его собственной личности (его недостаткам), но и указывает на некую поломку в научной системе. Как мы отмечали выше, она в значительной степени построена на доверии: по умолчанию каждый полагается на этичное поведение всех остальных. К сожалению, именно на подобной почве пышно цветет мошенничество — где всякие жулики, словно паразиты, могут удобненько выезжать на коллективном благорасположении научного сообщества. Бесстыдство действий Хвана лишь подчеркивает, насколько доверчивыми могут быть рецензенты и редакторы — те самые люди, на чей строгий скептицизм мы рассчитываем, — когда сталкиваются с такими потрясающими, "революционными" результатами.

Помимо фотографий клеток с микроскопа одной из самых распространенных мишеней для фальсификации изображений в биологии служат результаты блоттинга. Блоттинг — это методика, различные модификации которой используются специалистами по молекулярной биологии, чтобы разобраться, из чего состоит полученная ими в эксперименте

смесь. Исходно появился так называемый саузерн-блоттинг, названный в честь его изобретателя — биохимика Эдвина Саузерна[51]. С помощью саузерн-блотов детектируют последовательности ДНК: благодаря радиоактивным меткам получают изображения, которые вы, возможно, видели в статьях по генетике, сообщающих о новых результатах, — размытые полоски разной толщины, образующие своего рода вертикальные лесенки, или "дорожки"[52]. Другие виды блоттинга позволяют обнаружить РНК и белки и носят названия, сходные с первоначальным: нозерн-блоттинг, вестерн-блоттинг и так далее[53]. Биологические эксперименты зачастую основываются на проведении конкретного вида блоттинга, детектирующего конкретные соединения: например, вестерн-блот используется для диагностики некоторых заболеваний, выявляя выработку белков, свидетельствующих о присутствии определенных бактерий или вирусов. Ученые частенько с гордостью демонстрируют фотографии блотов в качестве иллюстраций к своим статьям — блот служит веским доказательством, что в эксперименте было обнаружено определенное вещество. Вот где раздолье для фальсификаторов.

Через десять лет после "открытий" Хвана, в 2014 году, ученые из японского института RIKEN опубликовали в журнале *Nature* две статьи, сообщающие о новых результатах по индуцированным плюрипотентным стволовым клеткам[54]. В отличие от стволовых клеток, замешанных в скандале с Хваном, *индуцированные* плюрипотентные стволовые клетки можно получить из уже зрелых клеток, то есть нет необходимости использовать именно материал эмбрионов[55]. Проблема в том, что стандартный процесс получения таких стволовых клеток, за который его изобретатели в 2012 году получили Нобелевскую премию, трудоемкий и малопроизводительный — он занимает несколько недель, и в результате получается много "мусора"[56]. Впрочем, группа из института RIKEN заявила, будто нашла другой способ получать такие клетки, придумав методику под названием STAP (*Stimulus-Triggered Acqui-*

sition of Pluripotency, “приобретение плюрипотентности под воздействием раздражителей”). Якобы все, что нужно сделать, — это подержать зрелые клетки в слабой кислоте (или подвергнуть их иному виду умеренного стресса, скажем воздействию давления), и тогда они превратятся в плюрипотентные стволовые клетки без каких-либо хлопот. Ведущий исследователь Харуко Обоката подобрала множество впечатляющих доказательств, включавших фотографии с микроскопа, диаграммы и результаты экспериментов с ДНК, которые свидетельствовали о том, что зрелые клетки перепрограммировались в плюрипотентные.

Открытие было революционным, и Обоката сразу обрела широкую известность в Японии. Пресса захлебывалась статьями о ней и ее чудаковатых порядках в лаборатории (домашний питомец — черепаха, лабораторные помещения украшены персонажами из мира муми-троллей, а вместо белого лабораторного халата она носит каппоги — японский фартук, подаренный ей бабушкой). Ее назвали ярким примером выдающейся женщины-ученого[57]. Продлилось это недолго. Через несколько дней после публикации статей Обокаты другие ученые начали замечать на рисунках несоответствия, особенно это касалось четырех дорожек на фотографии, запечатлевшей результаты электрофореза ДНК. Все четыре должны иметь одинаковый фон, раз они получены в одном и том же геле, однако при ближайшем рассмотрении видно, что фон одной из них темнее, к тому же имеет подозрительно четкие края. Выяснилось, что этот фрагмент был вставлен с фотографии другого геля и немного изменен в размерах для лучшего соответствия остальным дорожкам[58]. В статье об этом не говорилось ни слова — не очень-то похоже на действия ученого, работающего максимально прозрачно. После этой обнаружились и другие странности. Некоторые цвета на фотографиях были задним числом подкорректированы, а еще Обокату уличили в практике дублирования изображений: два якобы различных рисунка во второй ее ста-

тье были на самом деле одной и той же фотографией, только на одном рисунке она была — ничего неожиданного — зеркально отражена.

Тем временем, что довольно необычно, лаборатории по всему миру принялись лихорадочно пытаться воспроизвести эти результаты. Вероятно, одно из слабых мест методики STAP заключалось в ее поразительной простоте, так что другим ученым легко было все перепроверить. Один профессор по клеточной биологии создал сайт, куда исследователи могли посылать отчеты о выполняемых в режиме реального времени попытках повторить опубликованные результаты. Зеленый шрифт предназначался для положительных или обнадеживающих результатов, а красный — для провальных. По мере того как поступали отчеты, почти все они оказывались красными[59]. Под совокупным давлением проверяющих изображения и пытающихся повторить результаты институт RIKEN начал расследование, в ходе которого фальсификации с рисунками в статьях подтвердились. Обоката и ее коллеги обратились в редакцию журнала *Nature* с просьбой отозвать статьи, что и было сделано к июню 2014 года. А в декабре Обоката из RIKEN уволилась[60]. Более тщательное расследование выявило, что "послужной список" Обокаты содержит куда больше, чем исходные обвинения в подделывании изображений: она также включила в статьи рисунки из более ранних исследований, которые выдала за новые, и сфабрикованные данные, демонстрирующие, как быстро росли клетки. А любое "настоящее" доказательство плюрипотентности объяснялось тем, что она контаминировала образцы эмбриональными стволовыми клетками[61].

Финал истории с методикой STAP чудовищно печальный. Биолог Ёсики Сасаи, блестящий специалист по стволовым клеткам и соавтор в тех статьях, не замешанный в мошенничестве, но, согласно отчету RIKEN, несущий "тяжкую ответственность" за то, что не перепроверял результаты Обокаты, покончил с собой, повесившись в здании инсти-

тута в августе 2014-го[62]. Ему было пятьдесят два года. В своей предсмертной записке он упомянул негодование в средствах массовой информации, которое поднялось после раскрытия обмана Обокаты[63].

Истории Хвана и Обокаты обе явно необычны в одном отношении: мошеннические статьи снискали невероятную известность. Это были публикации в *Science* и в *Nature*, двух ведущих журналах мира. Довольно тревожно уже то, что две столь заметных пустышки сумели пройти процесс отбора в таких журналах, но еще престижность этих изданий означала, что статьи немедленно привлекут внимание всего мира — и внимание пристальное. Если такого рода мошенничество происходит на самом высоком научном уровне, получается, что куда больше фальсификаций остаются незамеченными, выходя в свет в менее известных журналах. Что ставит перед нами вопрос: как часто биологи подделывают изображения в своих статьях? В 2016 году микробиолог Элизабет Бик и ее коллеги решили это выяснить.

В сорока журналах по биологии они искали публикации, содержащие результаты вестерн-блоттинга, всего была отобрана двадцать тысяч шестьсот двадцать одна статья[64]. Поистине героически Бик лично просмотрела каждую публикацию, ища на фотоизображениях неправомерное дублирование. Она нашла достаточно, чтобы существенно пополнить галерею недобросовестных научных изображений: там было не только бесхитростное дублирование (см., к примеру, рисунок 1 на следующей странице), но и кадрирование в стиле Хвана, склеивание и подгон под нужные размеры в духе Обокаты и еще целый веер других нечестных приемов. В целом 3,8 % опубликованных статей (примерно одна из двадцати пяти) содержали сомнительное изображение. При последующем анализе публикаций из одного только журнала по клеточной биологии Бик с коллегами обнаружила еще большую долю — 6,1 %[65]. Во многих из этих статей просто содержались честные ошибки, и авторы могли решить проблему, опубликовав соответствующие

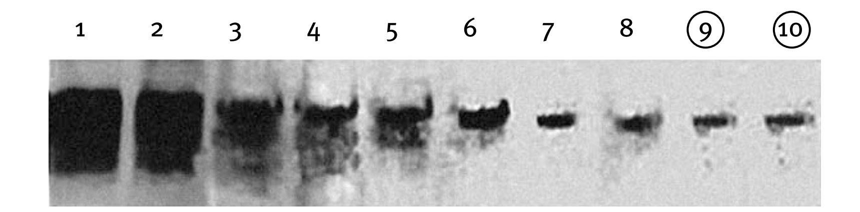


Рис. 1. Вестерн-блот с дублированием, найденный Бик и ее коллегами. Последние две полоски (дорожки 9 и 10) идентичны, изображение было задублировано (и то, что справа, слегка растянуто), вероятно, в программе вроде *Photoshop*. Та статья позднее была исправлена. Изображение взято из Bik E. M. et al., 2016.

поправки. Однако примерно 10 % тех статей были отозваны, а значит, в них содержалось что-то погнуснее. Если считать эти числа репрезентативными для статей по клеточной биологии вообще, то, по расчетам Бик, до тридцати пяти тысяч публикаций в научной литературе должны быть отозваны. Были и хоть какие-то хорошие новости: похоже, более престижные журналы в среднем реже публикуют статьи с дублированными рисунками. Пожалуй, самыми интригующими оказались результаты по рецидивистам: когда Бик и ее команда обнаруживали статью с поддельными рисунками, они проверяли, нет ли дублирования и в других публикациях того же автора. Оказывалось, что есть, без малого в 40 % случаев. Дублирование одного рисунка еще можно расценить как небрежность, дублирование двух выглядит мошенничеством.

Пока мы в основном говорили о жульничестве с изображениями, но мошенничают в науке далеко не только с ними. Вероятно, успешнее можно совершить — и спрятать — мошенничество с числами: строками и столбцами чисел, составляющих набор данных исследования. В предисловии мы познакомились с Дидериком Стапелом, который попросту вбивал нужные ему значения в таблицы и выдавал за реаль-

ные результаты. Как часто происходит подобное жульничество с данными? И насколько легко его вычислить?

По счастью, точно так же как необыкновенно сложно убедительно подделать Рембрандта или Вермеера (или убедительно сфабриковать фотографию вестерн-блота), столь же трудно фальсифицировать набор данных, не оставив улик. Данные, взятые с потолка, не обладают теми свойствами, каких мы ожидаем от данных, полученных в реальном мире[66]. По большому счету так происходит оттого, что никакая наука не может быть строго-строго точной: в числовых данных всегда есть *шум*. Всякий раз, как вы пытаетесь что-либо измерить, вы чуточку промахиваетесь мимо истинного значения, будь то экономические показатели страны, количество оставшихся в мире редких орангутанов, скорость субатомной частицы или даже просто рост человека. В случае с измерением роста, например, человек может немножко сгорбиться, мерная рулетка может сдвинуться на пару миллиметров или вы случайно запишете неверное значение. Это называют *погрешностью* (или *ошибкой*) *измерения*, и ее сложно полностью избежать, хоть и существуют способы ее уменьшить[67].

Столь же назойливым спутником исследований, как погрешность измерения, является *ошибка выборки*. Как ученые мы редко когда — если это вообще выполнимо — можем изучить все до единого проявления некоего феномена, неважно, исследуем ли мы совокупность клеток, или экзопланет, или хирургических операций, или финансовых сделок. Вместо этого мы берем выборки и пытаемся по ним сделать общие выводы для группы объектов в целом (статистики называют всю группу объектов “популяцией”, даже если речь идет не о группе людей). Беда в том, что характеристики любой сделанной выборки (скажем, средний рост всех людей в вашем исследовании) никогда не будут в точности тем, что вы действительно хотите узнать (допустим, средний рост всех жителей страны). В каждой выборке будет свое, чуть другое среднее, зависящее просто от того, кто по случайности в нее

попал. А для каких-то выборок, опять-таки по чистой случайности, значение может получиться сильно отличным от истинного среднего для группы в целом[68].

Как погрешность измерения, так и ошибка выборки непредсказуемы, но они предсказуемо непредсказуемы. Всегда можно ожидать, что данные по разным выборкам, измерениям или группам будут иметь несколько разные характеристики — в терминах среднего самое большое и самое маленькое значения, да и, в сущности, все остальное. Поэтому даже если обычно погрешность измерения и ошибка выборки — досадные помехи, они могут сослужить добрую службу, обеспечивая способ выявить сфабрикованные данные. Если набор данных выглядит слишком приглаженным и аккуратненьким, чересчур похожим для разных групп, тут наверняка кроется что-то странное. Как выразился генетик Джон Бёрдон Сандерсон Холдейн, "человек — животное упорядоченное", ему "крайне трудно имитировать разупорядоченность природы", и это относится к мошенникам в той же мере, что и ко всем остальным[69].

Подобные рассуждения позволили в 2011 году уличить социальных психологов Лоуренса Санну и Дирка Сместерса. Санна опубликовал работу, где утверждалось, что люди, находящиеся физически выше, оказались более просоциальными, а Сместерс заявлял, что якобы показал: когда люди видят красный и синий цвета, это влияет на их восприятие знаменитостей[70]. Результаты в обеих статьях на первый взгляд казались впечатляющими, однозначно подтверждая предложенные в них теории относительно человеческого поведения. Но при более тщательном изучении обнаруживалось нечто очень странное. Психолог Ури Симонсон выяснил, что для различных групп из эксперимента Санны разброс данных (разница между самым большим и самым маленьким значениями) практически идентичен, хотя группы вообще-то довольно сильно между собой различаются. Симонсон подсчитал: вероятность того, что подобное произойдет с ре-

альными данными, крайне низка. То же касалось и работы Сместерса, только у него слишком близкими оказались средние значения в группах, и опять же — такие похожие значения просто не могли получиться при работе с реальными данными, поскольку ошибки разбросали бы числа подальше друг от друга[71]. Как только эти проблемы — среди прочих — обнаружились, преступные статьи были отозваны, и оба исследователя с позором уволились[72]. Такого рода статистические красные флаги сродни сигналам опасности, вынуждающим банк заморозить вашу кредитку, после того как ее вдруг использовали для оплаты дорогущего тропического круиза: необычная активность, идущая вразрез с нормальными ожиданиями, может оказаться результатом мошеннических действий[73]. У сфабрикованных данных есть еще уйма других особенностей, способных заставить читателей статьи, когда те закапываются в детали, заподозрить неладное. К примеру, набор данных выглядит чересчур безукоризненно, в нем слишком мало выпавших точек, которые в реальных наборах данных появляются по всевозможным причинам: участники выбывают из исследования, приборы дают сбой и так далее. А может, распределение чисел не подчиняется ожидаемым математическим законам[74]. Или эффекты оказываются существенно больше тех, что мы находим правдоподобными в реальном мире, а значит, слишком хороши, чтобы быть правдой[75].

Некоторые фальсификаторы прекрасно знают, насколько трудно сделать так, чтобы поддельные числа выглядели правдиво, поэтому опробовали более творческие способы заметать следы. Политолог Майкл Лакур, на тот момент магистрант Калифорнийского университета в Лос-Анджелесе, в 2014 году опубликовал в журнале *Science* интересные результаты масштабного опроса на дому у респондентов[76]. Согласно полученным данным, если человека опрашивал гомосексуал, а не гетеросексуал, то это положительно влияло на мнение респондента об однополых браках, причем эффект был существенный и длительный. Таким образом, зна-

комство с представителем меньшинства, чьи права обсуждаются, сильно склоняло респондентов к поддержке этих прав. Оптимистичное заявление, и оно немедленно было использовано (в конечном итоге успешно) организаторами кампании за легализацию однополых браков в ходе ирландского референдума 2015 года[77].

Эти результаты впечатлили многих, в том числе двух других политологов, Дэвида Брукмана и Джошуа Каллу, которые захотели сами провести похожее исследование. Однако, изучая набор данных Лакура, они обнаружили крайне странные вещи. Они заметили, что распределение результатов, то есть разброс между высокими и низкими значениями на шкале, соответствующей степени одобрения однополых браков, подозрительно похож на разброс значений в другом, более раннем исследовании, хорошо им известном, — CCAP (*Cooperative Campaign Analysis Project*, "совместный проект по анализу исследований общественного мнения"). На самом деле распределения были почти идентичны. Данные последующих этапов исследования Лакура, когда тех же респондентов через некоторое время опрашивали повторно, чтобы проверить, не изменилось ли их мнение, тоже настораживали: ни один из участников не изменил своего мнения относительно первоначального больше, чем на некое определенное значение. Но ведь в числовых данных, как мы уже обсуждали, всегда есть шум: в подобных крупных наборах данных логично ожидать с течением времени куда больше колебаний в одну и другую сторону.

Данные Лакура оказались палимпсестом: он взял результаты исследования CCAP, немножко их подкрутил, добавив случайный шум, и выдал за результаты своей новой работы. Для последующих этапов он взял точно те же данные, еще немного их подкрутив, а многочисленные детали, описывавшие, как проходила подготовка опрашивающих, полностью выдумал. Исследование вообще не проводилось. Позднее его несчастный соавтор Дональд Грин, уважаемый профессор поли-

тологии, не замешанный в подделывании данных, поражался "невероятному количеству затейливейших выдумок с изощренными деталями", которые ему демонстрировал Лакур. "Там были всякие истории и случаи… графики и диаграммы. Казалось, никто не стал бы все это делать кроме как для того, чтобы изучить самый настоящий набор данных"[78].

Я отчетливо помню, как в мае 2015 года вышел отчет Брукмана и Каллы об исследовании Лакура[79]. Он появился в интернете как раз перед тем, как я сел в Эдинбурге на самолет, направляясь на конференцию в Сан-Франциско. К моменту посадки в США, через тринадцать часов, ученые, на которых я подписан в социальных сетях, только о нем и говорили. Отчет был разгромным, в нем в мельчайших подробностях изобличалась фальсификация Лакуром данных. Деваться ему было некуда: журнал *Science* вскоре отозвал статью, и Лакур распрощался с предложением работы, которое Принстонский университет сделал ему преимущественно из-за публикации в столь престижном журнале на самой заре карьеры[80].

Учитывая, сколько усилий Лакур затратил, дабы скрыть свой обман, ему определенно было бы проще выполнить исследование по-настоящему. Тогда бы он точно избежал негативных последствий в виде крушения своей карьеры, когда его набор данных изучили бы с подобающей тщательностью. Но Лакур, как Санна, Сместерс и Стапел до него, сфабриковал данные для того, чтобы все *контролировать*. Исследование удовлетворяло четким требованиям, предъявляемым рецензентами журнала *Science*, так что они точно приняли бы статью к публикации. Оно было именно тем, что требуют публикационная система и университетский рынок труда: не снимком беспорядочной реальности, где результаты нечетки, а интерпретации неопределенны, но ясным, убедительным результатом, который сразу же можно учитывать в реальной жизни.

Опять-таки эксплуатировалось не только желание рецензентов видеть интересные и важные открытия, но и их склон-

ность доверять. До *некоторой* степени доверие в процессе рецензирования неизбежно: рецензенты ведь не могут перепроверять каждую экспериментальную точку, выискивая признаки искажения результатов. Однако истории мошенничества с данными учат нас, что порог доверия может быть слишком низким, так что настоящий организованный скептицизм не срабатывает. Во имя науки ученым пора бы доверять друг другу чуть меньше.

Теперь мы готовы к вопросу: насколько мошенничество всех сортов и мастей распространено в науке? Один из способов оценить масштаб проблемы — посмотреть, сколько всего статей было отозвано. Отзы́в для статьи — это бесповоротный бесславный конец, его даже называют “высшей мерой наказания в науке”[81]. После такой смертной казни отозванные статьи попадают в своего рода чистилище. Они не просто отовсюду удаляются, поскольку это лишь вызвало бы еще большую путаницу, особенно если на них уже ссылались в других работах. Нет, они навечно остаются на сайте журнала с отметкой о том, что более не считаются легитимными. Часто эта отметка принимает форму слова “ОТОЗВАНА” (RETRACTED), напечатанного огромными жирными красными буквами и пересекающего по диагонали страницы статьи.

Лучше всего искать информацию об отозванных статьях на сайте *Retraction Watch* (“Наблюдение за отозванными статьями”), где отображается каждая новая отозванная работа, а у авторов и редакций журналов запрашивается комментарий, что пошло не так. В 2018 году владельцы сайта, Айвен Орански и Адам Маркус, создали базу данных, каталогизирующую более восемнадцати тысяч отозванных с 1970-х годов научных публикаций, — драгоценный клад для тех, кто интересуется скандальной стороной науки. Беглый просмотр столбца под названием “Причины отозвания” намекает на некоторые пикантные истории, стоящие за этими статьями: “конфликт ин-

тересов", "поддельное авторство", "неправомерные действия автора", "порча материалов и образцов", "судебное разбирательство"[82].

База данных *Retraction Watch* — список неидеальный: какие-то отозванные статьи могли быть пропущены, ведь не все журналы их открыто признают и заметно маркируют, издания очень сильно различаются в этом отношении. Еще важно отметить, что отзыв публикации не обязательно означает мошенничество — многие статьи аннулируются самими авторами, заметившими ошибку. Есть и более неоднозначные случаи: например, в начале 2020 года лауреат Нобелевской премии, химик-инженер Фрэнсис Арнольд объявила, что ее команда отзывает статью по ферментам из *Science*, поскольку результаты не воспроизведутся и "при тщательном изучении лабораторного журнала первого автора… выяснилось, что для ключевых экспериментов не хватает записей и сырых данных"[83]. Означало ли это простую ошибку, или на совести ведущего автора, студента из лаборатории Арнольд, было что похуже, неизвестно. Признание Арнольд было очень откровенным. "Прошу у всех прощения", — написала она в твиттере. "Я была немного занята, когда статью приняли к публикации, и не выполнила свою работу должным образом"[84].

Среди всех отозванных статей доля тех, которые содержат честные ошибки, составляет лишь 40 % или меньше. Преобладают те, что были отозваны из-за той или иной формы нарушения этики, включая мошенничество (около 20 %), повторное опубликование и плагиат[85]. Причем с течением времени число отозванных статей растет, хотя это и не обязательно подразумевает увеличение случаев мошенничества, скорее означает, что у редакторов журналов становится наметаннее глаз и авторы, как Арнольд, охотнее сознаются, что оплошали[86].

Так же как в обществе небольшое число нарушителей закона совершают несоразмерное количество преступлений, база данных *Retraction Watch* демонстрирует, что всего

лишь 2 % ученых ответственны за 25 % всех отозванных публикаций[87]. Худшие рецидивисты попадают в "список лидеров" *Retraction Watch*, это своего рода Нобелевская премия наоборот[88]. Наш знакомец Дидерик Стапел занимает там пятое место — с пятьюдесятью восемью отозванными статьями*. Правда, звание абсолютного чемпиона-тяжеловеса сейчас занимает японский анестезиолог Ёситака Фудзии, который выдумал данные для несуществовавших испытаний лекарства и чье количество отозванных публикаций достигает невероятного числа — ста восьмидесяти трех**. В 2000 году в письме редакторов журнала *Anaesthesia & Analgesia* опубликованные Фудзии данные описывались как "поразительно красивые"[89]. Как отметили менеджеры сайта *Retraction Watch*, это был не комплимент[90]. Авторы того письма обратили внимание, что в испытаниях лекарства у Фудзии число участников, сообщивших о головной боли в качестве побочного эффекта, совершенно идентично для различных групп из тринадцати его исследований и почти идентично для еще восьми. Данные, так же как в случаях, рассмотренных нами выше, были распределены слишком уж равномерно, чтобы походить на правду. Однако больше десяти лет не предпринималось абсолютно ничего, и все это время Фудзии продолжал клепать поддельные статьи в уважаемые журналы по анестезиологии. И только в 2012 году в ходе другого анализа обнаружилось, что еще больше его данных чрезвычайно неправдоподобны, и формальное расследование положило конец его карьере[91]. В дополнение к подтвержденному списку из ста семидесяти двух статей, которые содержали поддельные данные (с тех пор список пополнился, что вывело Фудзии на позицию рекордсмена), исследователи перечислили и те его публикации, что, по их мнению, *не* содержали никакого обмана. Таких было три[92].

* Список с тех пор изменился, теперь Стапел там на седьмом месте.

** Его уже успели обойти: по состоянию на июль 2023 года Фудзии занимает второе место со ста семьюдесятью двумя статьями.

Итак, если объединить тяжеловесов из “списка лидеров” *Retraction Watch* со всеми остальными исследователями, чьи статьи оказывались отозванными по нечестным причинам, сколько же ученых на самом деле мошенничают? Совокупная доля отозванных статей — примерно четыре на десять тысяч публикаций, то есть 0,04 % — ободряюще мала. Это не очень-то помогает, ведь мы знаем, что, с одной стороны, часть статей отзывается не из-за мошенничества, а с другой — некоторые журналы либо не отлавливают фальшивые результаты, либо не утруждаются тем, чтобы соответствующие публикации отзывать. А что будет, если мы просто спросим ученых, анонимно, мошенничали ли они когда-нибудь?

Самое масштабное на сегодняшний день подобное исследование объединило результаты семи опросов: оказалось, 1,97 % ученых признают, что хотя бы однажды подделывали данные[93]. Один ученый, сознавшийся в мошенничестве, на пятьдесят — это и так достаточно тревожный показатель, однако примите во внимание, что люди обычно неохотно признаются в обмане, даже и в анонимном опросе, так что реальное число точно гораздо больше. И действительно, когда ученых спрашивали, сколько они знают *других* исследователей, которые фальсифицировали данные, значение подскакивало до 14,1 % (хотя, конечно, кто-то из опрошенных мог ошибаться, преувеличивать проблемы с чужими исследованиями или напрасно подозревать своих конкурентов)[94].

Кто же эти жулики? Возможно ли составить “профайл”, как делают в ФБР, типичного мошенника, психологический портрет, который помог бы нам предотвращать новые акты подделывания данных? В обзоре случаев мошенничества специалист по нейронаукам Чарльз Гросс сокрушался, что нет четких критериев, по которым можно было бы понять, кто обманывает. Однако он попытался описать тип человека, регулярно появляющегося в средствах массовой информации в сюжетах о мошенничестве, получающих широкую огласку. Как правило, писал Гросс, это “яркий и амбициозный моло-

дой человек, работающий в элитарном институте в быстро развивающейся и высококонкурентной области современной биологии или медицины, где результаты имеют важное теоретическое или клиническое применение либо финансовые последствия"[95]. Привычная картина, если учесть, что мы обсуждали в этой главе: например, Паоло Маккиарини практически идеально подходит под данное описание.

Примечательно, что Гросс описал мошенника как мужчину. Это четко прослеживается для самых закоренелых обманщиков: из тридцати двух ученых, составляющих сейчас "список лидеров" *Retraction Watch,* женщина только одна*[96]. Чтобы понять, говорит ли нам это о чем-то важном, надо знать, каким было типичное соотношение мужчин и женщин в каждой из рассматриваемых областей — и следовательно, не были ли мужчины перепредставлены. В исследовании 2013 года, фокусировавшемся на науках о жизни, это исходное различие учитывалось — и выяснилось, что мужчины действительно перепредставлены среди героев сообщений о мошенничестве от Службы по обеспечению добросовестности в научной практике США (ORI)[97]. А в исследовании 2015 года, в ходе которого изучались отозванные и исправленные публикации во всех научных областях, не было обнаружено гендерных различий, хотя неясно, учли ли в нем важнейшее исходное соотношение[98].

Собрав свою базу данных из статей с дублированными изображениями вестерн-блотов, Элизабет Бик и ее коллеги также проверили, нет ли каких-то особенностей, отличающих проблемные статьи от остальных[99]. Выделялось одно обстоятельство: дублирование изображений случалось в некоторых странах чаще, чем в других. Так, Индия и Китай были перепредставлены, тогда как США, Великобритания, Германия, Япония и Австралия — недопредставлены. Авторы предположили, что это связано с культурными различиями: в странах

* Сейчас в списке тридцать один ученый, но соотношение мужчин и женщин примерно такое же.

вроде Индии и Китая из-за более гибких правил и более мягких наказаний за мошенничество в науке проводится больше потенциально жульнических исследований[100]. Что еще раз подчеркивает: социальная среда, в которой делается наука, может серьезно влиять на качество исследований.

К похожим выводам приходили и другие люди, изучавшие этот вопрос. Сославшись на статью, в которой говорилось о том, что довольно подозрительным образом *сто процентов* клинических исследований акупунктуры, проводимых китайскими учеными, дают положительные результаты (даже если бы акупунктура прекрасно работала, логично было бы ожидать каких-то случайных отрицательных результатов), врач и писатель Стивен Новелла выразил мысль, что политические обстоятельства в Китае, вероятно, не благоприятствуют высококачественной науке:

> Существует также обоснованное опасение, что в тоталитарных государствах не создается среда, в которой наука может процветать. Наука требует прозрачности, она требует ценить методы выше результатов, и она должна быть идеологически нейтральна. А это не те идеи, что успешно развиваются при тоталитарных режимах. Кроме того, ученые, которых назначают на какие-то высокие, влиятельные должности, — это чаще всего те, кто угождает режиму, доказывая, например, что культурная пропаганда правдива. Таким образом, давление отбора, связанное с продвижением по службе, не способствует добросовестности в науке[101].

Китайские ученые, наверное, согласились бы, что серьезная проблема есть, каковы бы ни были ее причины. В ходе одного опроса в начале 2010-х годов китайские биомедицинские исследователи оценили количество статей в данной области, пораженных каким-либо видом нарушения научной этики, примерно в 40 % от всех биомедицинских публикаций, написанных их соотечественниками; 71 % участников сообщил, что

власти в Китае “совсем или почти не обращают внимания” на случаи мошенничества[102].

Впрочем, не считая общих соображений о гендере или стране происхождения, мы очень расплывчато представляем себе, каковы особенности типичных мошенников от науки. Если мы не в силах выявить их по демографическим признакам, поможет ли нам понимание их мотивов? Зачем мошенники совершают столь грубые надругательства над наукой, особенно когда им очень даже есть что терять? В исследовании 2014 года обнаружилось, что ученые в США, которым Служба по обеспечению добросовестности в научной практике вынесла осуждение за нарушения научной этики, в предшествовавшие годы испытывали большие трудности с финансированием. Это может означать, что нужда в грантовых средствах играет свою роль, однако есть и альтернативное объяснение: что финансирование иссякло, пока в отношении этих ученых велось расследование (то есть скорее нарушения научной этики привели к недостатку финансирования, чем наоборот)[103].

Еще одна причина может заключаться в патологически ошибочном представлении мошенника о том, что такое наука. Иммунолог и нобелевский лауреат сэр Питер Медавар утверждал, что, как ни странно, ученые, совершающие мошенничество, *слишком сильно* радеют об истине, только вот их представление о том, что есть истина, оторвано от реальности. Он писал: “Я думаю, что самое важное побуждение к мошенничеству в науке — страстная вера в правду и значимость теории или гипотезы, которую не принимают во внимание или открыто не признают большинство ученых-коллег, коих нужно, соответственно, заставить признать то, что, по мнению уязвленного ученого, является самоочевидной истиной”[104]. Физик Дэвид Гудстайн согласен: он полагает, что “введение лжи в тело науки крайне редко бывает целью тех, кто совершает мошенничество”. “Они почти всегда верят, что вводят в научную летопись правду… просто не пре-

одолевая все те трудности, что предписываются настоящим научным методом"[105].

Отчасти Гудстайн говорил об известном примере из своей собственной области. В 2001 году немецкий специалист по физике конденсированного состояния Ян Хендрик Шён, работавший в прославленных "Лабораториях Белла" в США, поразил весь мир, заявив, что создал основанный на углероде транзистор (это устройство управляет электрическими сигналами, переключая и усиливая ток, и служит основополагающим элементом фактически любой электрической цепи), в котором переключение происходит на уровне одной молекулы[106]. Его транзистор был куда компактнее всех тех, что можно сделать из кремния, стандартного материала для микрочипов, и сулил коренные перемены в области построения электрических цепей. В конце концов мы начали бы строить электрические цепи молекулярного масштаба, что дало бы мощнейший толчок развитию нанотехнологий. Стэнфордский профессор восхитился "чрезвычайно элегантной в своей простоте" технологией, и Шён получил многочисленные награды от разных научных обществ[107]. Микроскопический транзистор и быстрый технический прогресс обеспечили Шёну абсолютное первенство по публикационной активности: между 2000 и 2002 годом вдобавок ко множеству статей в превосходных физических журналах он ухитрился опубликовать девять статей в *Science* и семь — в *Nature*. Для большинства ученых даже одна статья в одном из этих сверхсолидных журналов стала бы пиком карьеры. Поговаривали о Нобелевской премии.

Однако вскоре о Шёне стали перешептываться с подозрением. Другие лаборатории, пытаясь повторить его эксперименты (они довольно просты, как и в случае с методикой STAP для стволовых клеток, так что несложно было все перепроверить), столкнулись с колоссальными трудностями. Затем обнаружилось, что сразу в нескольких статьях Шёна, якобы описывавших совершенно разные эксперименты, по-

казан один и тот же иллюстрирующий результаты рисунок[108]. "Лаборатории Белла" начали тщательное расследование, попросив у Шёна сырые данные, подтверждающие его открытия. Но "его домашнее задание съела собака": Шён сказал, что удалил бóльшую часть своих данных, поскольку у его компьютера "не хватало памяти"[109]. В тех данных, что он все-таки предоставил, расследовавший его дело комитет все равно сумел найти четкие доказательства подлога. Например, при якобы проведенном сравнении токов для двух разных молекул он просто удвоил все значения для первой молекулы и выдал их за результат для второй. Если разделить значения для второй молекулы на два, получатся те же числа, что и для первой, — вплоть до пятого знака после запятой[110]. В ходе расследования всплыло и много других примеров дублирования, подтасовок и полноценного подделывания данных. И снова то же: наглый мошенник навыдумывал данных, которые слишком аккуратны, чтобы быть реальными.

Впрочем, для нас как для охотников за мотивами мошенников особенный интерес представляет концовка длинного отчета комитета по расследованию — краткий ответ самого Шёна. Он заявляет, что, хотя и "совершал ошибки" и "осознает, что [результаты] недостаточно надежны", он также "искренне верит, что опубликованные им научные результаты реальны, волнующи и стоят того, чтобы ради них работать"[111]. Конечно, мы должны быть крайне осмотрительны и не должны верить обманщикам на слово. Но очевидная вера Шёна в свою теорию даже после того, как значительная часть его работы была вычеркнута из научной литературы (сейчас он на пятнадцатом месте в "списке лидеров" *Retraction Watch* с тридцатью двумя отозванными статьями*), говорит о своего рода делюзии и самообмане, упоминавшихся Медаваром и Гудстайном[112]. Мы не можем знать наверняка, но Шён, вероятно, искренне верил в свои расчудесные транзисторы и счи-

* Уже на двадцать втором месте.

тал, что в стремлении привлечь к ним внимание всего мира нарушение правил — это необходимое зло.

Дидерик Стапел в своей книге выразил похожую мысль, написав, что начал фальсифицировать статьи по социальной психологии после того, как стал получать разочаровывающие результаты в тех исследованиях, что должны были, по его мнению, увенчаться успехом:

> Когда результаты не вполне те, на которые ты так отчаянно надеялся, когда ты знаешь, что надежда эта основана на тщательном анализе литературы, когда это твой третий эксперимент на данную тему, а первые два получились прекрасно, когда ты знаешь, что где-то и другие люди выполняют похожее исследование и все у них идет хорошо, тогда ты, наверное, имеешь право чуточку подправить результаты?[113]

Таким образом, мошенники вроде Шёна и Стапела представляют собой экстремальный вариант явления, с которым мы будем сталкиваться в этой книге снова и снова: ученых, позволяющих себе протянуть руку помощи "истине" — или тому, что они *хотят* видеть истиной[114].

Каковы бы ни были причины, мошенники от науки наносят ей сокрушительный и несоразмерный вред, а значит, вредят одному из самых ценных общественных институтов человечества. Во-первых, учтите потери времени. Расследование мошеннического случая может занять недели, месяцы или даже годы интенсивной работы, потому, в частности, что нарушения научной этики зачастую метастазируют, проникая из исходного исследования во многие другие, и каждое нужно проверить с криминалистической дотошностью. Большинство людей, выполняющих такую работу, не занимаются только ею одной: часто это самостоятельные исследователи, ставящие собственные проекты на паузу, чтобы во имя

общего блага расследовать обман. К тому же процесс на сборе доказательств не заканчивается. Как подтвердили бы анонимные разоблачители, расследователи в мире данных и любые другие люди, кто контактировал с научным журналом или институтом, уличенным в каких-либо нарушениях, отзыв даже откровенно мошеннической статьи — процесс медленный, если вас вообще попросту не проигнорировали или сразу не отфутболили.

К временным затратам, разумеется, добавляются и финансовые. Потери от непосредственной кражи, как в том случае, когда Хван У Сок присвоил себе часть грантовых средств, меркнут в сравнении с потерей всех тех денег, что тратятся на мартышкин труд — на исследования, которые мы тщимся выстроить на результатах, коих никогда и не было. Например, исследователь ожирения Эрик Поэлман, первый мошенник от науки, которого посадили в тюрьму в Америке, выбросил на ветер миллионы долларов налогоплательщиков в виде грантов от правительства США, десять лет штампуя бесполезнейшие поддельные данные[115]. А сколько абсолютно невинных ученых растратили грантовые деньги, выделенные на их собственные исследования, пытаясь развить или воспроизвести работу Поэлмана либо других мошенников?

Вдобавок к бессмысленным тратам времени и денег мошенничество ужасно деморализующе действует на ученых. Как мы видели, одна из причин, по которым так много вранья ухитряется проникнуть в литературу, состоит в том, что в целом ученые — люди открытые и склонны доверять. Для рецензентов привычное дело — скептично следить за тем, как результаты интерпретируются, но мысль, что данные могут быть поддельными, им обычно и в голову не приходит. Однако явная распространенность мошенничества означает, что всем нам нужно добавить угнетающую опцию в свой репертуар реакций на подозрительно выглядящие статьи: нам могут лгать. И с этой особой бдительностью надо рассматривать не только чужие статьи — мошенничество может происходить

прямо под носом у любого ученого. Публикации редко выходят за авторством одного-единственного исследователя, поэтому нечистый на руку соавтор иногда может испортить репутацию целой команды невинных коллег. Во многих случаях виновник — младший (по чину) сотрудник лаборатории, который позорит имена своих старших соавторов, как в истории Майкла Лакура с выдуманным опросом мнений насчет однополых браков. Бывает и наоборот: состоявшиеся ученые лихо ставят на кон карьеры своих подчиненных (в отчете о мошенничестве Дидерика Стапела, например, говорилось, что не менее чем у десяти его аспирантов темы диссертаций основывались на его поддельных данных)[116]. И мы уже видели, чего может стоить крах репутации, как в случае Ёсики Сасаи, который лишил себя жизни, когда понял, что втянут в скандал вокруг методики STAP, связанной со стволовыми клетками.

Еще мошенничество загрязняет научную литературу. Хотя отзыв статьи заметно уменьшает количество ее цитирований, часто этого недостаточно: на отозванные публикации все равно регулярно ссылаются рассеянные ученые[117]. В 2015 году исследователи проследили историю мошенника-анестезиолога Скотта Рубена, у которого к 2009 году было отозвано двадцать статей (сейчас он на двадцать седьмом месте в "списке лидеров" *Retraction Watch* с двадцатью четырьмя отозванными публикациями*). В последующие пять лет на эти статьи суммарно сослались двести семьдесят четыре раза, и лишь примерно в четверти случаев ссылающиеся ученые, по-видимому, понимали, что публикации уже отозваны[118]. Анализ нескольких других отозванных статей показал, что 83 % цитирований после отзыва были положительного, утвердительного характера и не упоминали факт отзыва — эти публикации-зомби по-прежнему шевелились в научной литературе, и мало кто замечал, что они мертвы[119]. Некоторые журналы очевидно хуже других помечают отозванные статьи[120].

* Его обошли другие плодовитые мошенники, в "список лидеров" он уже не входит.

Независимо от причин то обстоятельство, что на отозванные статьи продолжают регулярно ссылаться, причем не ради критики, означает следующее: громадное количество научных исследований основывается на совершенно ложной информации. Акт мошенничества затрагивает не просто исходную статью, где подаются поддельные результаты, а куда большую часть научной литературы.

Урон наносится не только науке: влияние мошенничества расползается далеко за пределы журналов, где публикуются фальшивки. Специалисты-практики, полагающиеся на научные исследования, например врачи, могут быть введены в заблуждение и в результате использовать средства лечения или методики, которые не работают либо же откровенно опасны. Пример последнего — случай анестезиолога Йоахима Больдта, номера два в “списке лидеров” *Retraction Watch*[*] [121]. Больдт сфабриковал данные по гидроксиэтилированному крахмалу — химическому соединению, иногда применявшемуся при операциях в травматологии для восполнения объема крови (считалось, что оно может предотвратить шок при большой кровопотере, помогая оставшейся крови циркулировать по сосудам). По подделанным Больдтом результатам получалось, что якобы гидроксиэтилированный крахмал для применения в этих целях безопасен. Такой вывод подкреплялся тем обстоятельством, что метаанализ — обзорное исследование, объединяющее все предшествующие статьи на данную тему, — дал то же заключение. Однако так вышло только потому, что Больдта тогда еще не разоблачили, так что в метаанализе учитывались его поддельные результаты. Когда же об обмане Больдта стало известно и его статьи из метаанализа исключили, результаты кардинально поменялись: на самом деле пациенты, которым вводили гидроксиэтилированный крахмал, умирали чаще[122]. Мошенничество Больдта исковеркало целую область исследований, поставив

* С тех пор он вырвался вперед — на первое место.

под угрозу тех пациентов, чьи хирурги, что совершенно естественно, приняли его результаты за чистую монету[123].

Один из самых ужасных случаев мошенничества в науке не только сбил с толку ученых и врачей, но еще и сильнейшим образом повлиял на восприятие людьми жизненно важного медицинского средства. Он вызвал столько страха и хаоса, что, хотя произошло это более двух десятков лет назад, мы по сей день ощущаем его вредное воздействие. Я говорю, конечно же, о печально известном исследовании вакцин, результаты которого были опубликованы в 1998 году в журнале *The Lancet* британским врачом Эндрю Уэйкфилдом[124]. На основании выборки из двенадцати детей Уэйкфилд и его соавторы сделали вывод, что комбинированная вакцина от кори, паротита и краснухи (MMR — *measles*, *mumps*, *rubella*) связана с аутизмом. Якобы вирус кори, остающийся в организме после прививки, вызывал симптомы аутизма в кишечнике и мозге (Уэйкфилд назвал это состояние, описываемое им впервые, "аутистическим энтероколитом")[125]. В интервью и на пресс-конференции после выхода статьи Уэйкфилд неоднократно заявлял, что MMR нужно разбить на три отдельных вакцины, поскольку с их комбинацией "иммунная система некоторых детей справиться не в силах"[126].

Сейчас уже большинство людей знают, что результаты Уэйкфилда опровергнуты. С 1998 года проводилось несколько масштабных строгих исследований, в ходе которых не было выявлено никакой связи между вакциной MMR (и любой другой, кстати) и расстройствами аутистического спектра[127]. Еще было показано, что комбинированные вакцины так же безопасны, как и некомбинированные[128]. А вот чего многие *не* знают: статья Уэйкфилда была далеко не честной ошибкой или же вполне объяснимым тупиком в предварительном, пристрелочном направлении исследований, а мошенничеством от начала и до конца[129].

После выхода публикации и возникновения разногласий журналист-расследователь Брайан Дир начал изучать данные

Уэйкфилда и, самое главное, его побуждения. В серии шокирующих статей, появившихся в журнале *The BMJ* (ранее *British Medical Journal*), Дир описал, как Уэйкфилд исказил или подделал медицинские сведения о каждом ребенке из тех двенадцати, что составляли его выборку[130]. Он попросту выдумал тот "факт", что у всех детей первые симптомы, связанные с аутизмом, проявились вскоре после вакцинации препаратом MMR, тогда как на самом деле у кого-то симптомы были зафиксированы уже до того, у кого-то проявились только через несколько месяцев, а некоторым детям вообще даже никогда и не ставили диагноз "аутизм"[131]. Что же касается побуждений, Дир выяснил, что у Уэйкфилда было два существенных финансовых мотива повернуть исследование так, как он его и повернул[132]. Во-первых, один адвокат платил ему солидные деньги, поскольку собирался от лица родителей тех детей с аутизмом подать в суд на производителей вакцины[133]. Более того, Уэйкфилд именно через связанную с этим адвокатом антивакцинаторскую группу и находил пациентов для своего исследования. Во-вторых, за год до публикации результатов он подал заявку на патентование своей собственной вакцины от кори, и поэтому ему выгодно было, запугав людей, отвратить их от комбинированной вакцины MMR[134]. Непростительным образом ни один из этих "конфликтов интересов" не был указан в статье: в ней просто отмечалось, что работа финансировалась "особыми попечителями" и родители детей "дали стимул для проведения этих исследований"[135].

Когда Дир впервые поделился с редакцией журнала *The Lancet* своими соображениями, а было это в 2004 году, он столкнулся с ожесточенным сопротивлением (что, надо заметить, предвосхитило ту же реакцию того же журнала в случае с Маккиарини примерно десятью годами позже)[136]. Из-за расследования Дира статья Уэйкфилда в итоге была в 2010 году отозвана, пробыв к тому моменту частью официальной научной литературы двенадцать лет. Уэйкфилда также вычеркнули из реестра Главного медицинского совета Великобритании,

после самых длительных слушаний в истории организации, лишив его права работать в этой стране врачом. И не только за фальсификацию данных, но еще за то, что он подверг ребенка ненужным медицинским процедурам, в том числе колоноскопии, без надлежащего одобрения[137]. Совет охарактеризовал его как человека, своими действиями проявляющего "черствость и неуважение", — меткое описание, подходящее любому мошеннику из этой главы[138]. С тех пор Уэйкфилд прославился как участник антипрививочного движения в США. Так, в 2016 году его антивакцинаторский фильм *Vaxxed* едва не показали на кинофестивале "Трайбека", проходящем в Нью-Йорке, — фильм поддерживал основатель фестиваля и "вакцинный скептик" Роберт Де Ниро, однако под давлением общественности снял его с программного показа[139].

Обман Уэйкфилда породил страх перед вакцинами, который распространялся столь же быстро, словно какой-нибудь вирус. Во многих разделах британской прессы начали выходить статьи о вакцине MMR в стиле "просто спросить", сея сомнения среди родителей. Зачинщицей стала газета *Daily Mail*, и даже журнал *Private Eye*, который гордится тем, что не ведется на мнение большинства, посвятил вакцине MMR в 2002 году специальный номер, выведя Уэйкфилда фигурой масштаба Галилея, противостоящей влиятельным кругам, которые продвигают комбинированную вакцину[140]. Столь широкое освещение в средствах массовой информации возымело именно тот эффект, какого и следовало ожидать[141]. В конце 1990-х годов в Великобритании показатель принятия вакцины MMR обществом приближался к 95 %, необходимым для достижения так называемого коллективного иммунитета — когда привито достаточно людей, так что заболевание становится довольно редким и не угрожает тем, кто по медицинским причинам, например из-за аллергии на компоненты препарата, не может от этой хвори вакцинироваться. А вот после 1998 года этот показатель рухнул аж до 80 %, а показатели заболеваемости корью, соответственно, подскочили[142].

В Европе и за ее пределами начались эпидемические вспышки, и в странах, уже какое-то время считавшихся свободными от кори, стали регистрировать новые случаи. Согласно последней оценке Всемирной организации здравоохранения, в 2018 году свыше ста сорока тысяч человек умерло от кори и ее осложнений, что особенно трагично и обидно, поскольку это болезнь, которую мы умеем предотвращать[143]. Не будет большим преувеличением сказать, что из-за научного мошенничества Уэйкфилда мир теперь гораздо опаснее — особенно для детей и других уязвимых людей, тем более в развивающихся странах.

Принятие к публикации статьи Уэйкфилда таким видным журналом, как *The Lancet*, станет восприниматься как одно из худших решений в истории научных изданий. Едва ли удалось бы найти более наглядный пример того, насколько надежная наука важна для благополучия общества и как система рецензирования не справилась с отсеиванием плохого исследования. Это снова возвращает нас к вопросу доверия — на сей раз доверия общества к науке. Вакцинировать своего ребенка — значит совершить активное действие, веря, что медики, считающие это безопасным, правы[144]. Если исследование в известном журнале, одобренное научными рецензентами, дает понять, что все не так, в высшей степени разумно будет принять это к сведению. Годами после той публикации люди действительно не знали, кому верить насчет вакцин. Многие не знают до сих пор[145].

Злоупотребление общественным доверием — вероятно, самый губительный аспект научного мошенничества. Люди вложили в науку очень много, включая здоровье своих детей. Мошенники же сводят это доверие на нет. Хотя в первую очередь винить, безусловно, нужно самих обманщиков, наша научная система тоже заслуживает порицания. Мало того, что самые знаменитые журналы вынуждают ученых по-

сылать им только эффектные результаты — тем самым, в общем-то, гарантируя, что кто-то примется жульничать, дабы достичь требуемой эффектности, — часто редакторы журналов действуют с неохотой и явным сопротивлением, когда всплывают даже вполне веские доказательства каких-то нарушений[146]. Институтам тоже не избежать критики: мало того, что их расследования часто довольно вялые, мы видели случаи вроде истории Маккиарини, когда институты защищали непростительное и активно преследовали разоблачителей[147]. Естественно, человек, обвиненный в мошенничестве, должен считаться невиновным, пока не доказано обратное; лучший способ разрушить доверие — разбрасываться голословными обвинениями. Но чем дольше те, кто занимает ответственные должности, тянут резину и чем дольше каждая фальшивая статья присутствует в научной литературе, тем больше наши структуры и учреждения подводят науку — а в конечном счете и общество.

И словно от всех историй о загрязнении научной литературы, обо всех потерях, об изъязвлении доверия и даже о смертях нам еще недостаточно скверно, снова всплывает пугающая мысль: *это лишь те случаи, о которых мы знаем*. А вдруг мошенники, которых мы не поймали, умнее, хитрее и опаснее, чем все обсуждавшиеся в этой главе? В конце концов, многих мошенников от науки выводят на чистую воду не расследователи в мире данных или те, кто ищет в статьях поддельные изображения, а разоблачители, просто оказавшиеся в нужном месте в нужное время, чтобы заметить нечто подозрительное. Наверняка есть мошенники, избежавшие обнаружения, кто скрыл свои злодеяния более умело и состряпал поддельные научные результаты, не вызывающие никаких вопросов. Не исключено, что их мы никогда не найдем[148].

Зачем начинать погружение в многочисленные проблемы науки с мошенничества? Не должны ли эти истории, зачастую

столь драматичные и огорчительные, быть кульминацией нашего путешествия в наихудшее из предлагаемого наукой? На первый взгляд это кажется естественным. Но какими бы ужасными эти случаи мошенничества ни были, то, что я буду описывать в следующей главе, в определенном смысле гораздо хуже. И хуже именно потому, что *не так* очевидно и заметно, как поддельные изображения или фальшивые данные. В отличие от мошенничества, представляющего собой явное, непростительное нарушение законов, по которым ведется научная деятельность, предмет нашего дальнейшего обсуждения имеет отношение к хорошим и честным научным устремлениям. Это проблема куда более тонкая, незримая и, что всего хуже, гораздо более распространенная.

Глава 4
Предвзятость

> Однажды принятая гипотеза придает нам рысью зоркость по отношению ко всему, что ее подтверждает, и делает нас слепыми по отношению ко всему, что ей противоречит*.
>
> Артур Шопенгауэр *"Мир как воля и представление" (1818)*[1]

> Наука... совершит самоубийство, если признает какие-то догмы.
>
> Томас Гексли *"Памятник Дарвину" (1885)*[2]

В 1830–1840-х годах Сэмюэль Мортон, знаменитый американский врач и ученый, опубликовал серию богато иллюстрированных книг, содержавших его измерения сотен человеческих черепов со всего мира[3]. Его метод заключался в том, чтобы заполнить все пустоты черепа горчичными семенами (позднее — свинцовой дробью), а затем оценить, насколько крупным должен был быть мозг внутри черепа, по количеству семян или дробинок, которые удалось затолкать внутрь[4]. По результатам измерений своей коллекции он заключил, что черепа европейцев вместительнее, чем черепа азиатов, индейцев и африканцев, и предположил, что эти различия демонстрируют отличные для разных групп "умственные и нравственные способности"[5]. Книги Мортона, в которых он также обсуждал свои надуманные теории о совершенно различном происхождении разных человеческих рас, стали международной сенсацией и сыграли ключевую роль в укреплении позиций расовой теории, пытавшейся разделить людей на иерархию высших и низших групп и способ-

* Перевод Ю. Айхенвальда. Цит. по: Шопенгауэр А. *Мир как воля и представление* // Собр. соч. в 6-ти т. Т. 2. М.: Республика, 2001.

ствовавшей разжиганию некоторых из наижутчайших кошмаров XIX и XX столетий.

Наряду с различиями средних значений по группам Мортон предоставил изобильные данные по своим измерениям большинства черепов. Подобная степень прозрачности для того времени была нетипична и позволила будущим исследователям перепроверить его данные. В 1978 году, когда Мортона и его теории уже почти полностью позабыли, палеонтолог Стивен Джей Гулд именно это и сделал.

Мортоновский анализ черепов, писал Гулд, страдал от целого ряда неувязок. Мортон делил группы произвольно: например, сообщал результаты для некоторых подгрупп черепов белых, где каждая характеризовалась высокими средними, но не делал того же для некоторых подгрупп индейцев, тоже обладавших крупными черепами. В некоторые группы он несправедливо включал больше мужчин — а ведь у них, как мы знаем, головы больше, поскольку и тела вообще крупнее, — тем самым неправомерно увеличивая среднее. Вычисления для каких-то групп он перепроверил на предмет ошибок, а для других — нет. Еще между измерениями с помощью семян и измерениями с помощью более надежных свинцовых дробинок было расхождение — и для черепов черных и индейцев оно было больше, чем для белых, а значит, ошибки в измерениях с помощью семян происходили избирательно. Позднее Гулд предложил "вероятный сценарий" того, как это могло происходить:

> Мортон, проводя измерения с помощью семян, берет угрожающе большой череп черного, небрежно его заполняет и несколько раз легонько встряхивает. Затем берет удручающе мелкий череп кавказца, усердно его трясет и сильно нажимает большим пальцем* в районе большого затылочного отверстия

* Гулд объясняет: семена, в отличие от дробинок, легкие и разные по размеру, поэтому упаковываются неплотно. Если хорошенько трясти череп или нажимать пальцем на большое затылочное отверстие, это способствует тому, что семена утрясываются — и освобождается место для новых.

> [находится в основании черепа, через это отверстие входит позвоночник]. Это происходит легко, без сознательных побуждений; ожидания — мощное руководство к действию[6].

Таким образом у Мортона получалось бы, что черепа популяций белых крупнее, чем небелых. И действительно, все его огрехи смещали результаты в том же направлении. Ошибки, как выразился Гулд, отражали "тиранию априорного предпочтения", то есть убеждение Мортона в превосходстве белых[7]. Если анализировать данные должным образом, то различия между черепами разных этнических групп оказывались бы совсем незначительными — уж точно не на чем было бы выстраивать расовую иерархию. И это не единичная история. Те же уроки о влиянии предвзятости, говорил Гулд, вероятно, применимы ко всей науке: "Подозреваю, что неосознаваемые или слабо ощущаемые жульничество, искажение и манипулирование данными широко распространены, повальны и неотвратимы в профессии, присваивающей статус и авторитет за аккуратные и недвусмысленные открытия"[8].

Гулд был абсолютно прав. Написал он эти слова в 1970-х, и с тех пор стало еще очевиднее: ученые регулярно проводят свои исследования таким образом, что до сознательного мошенничества не доходят, однако же сильно сдвигают шансы в свою пользу. И хотя далее мы вернемся к идеологическим предубеждениям ученых, политические взгляды, включая те, что подобны расовым предрассудкам, приписанным Гулдом Мортону, не являются главной темой этой главы. Прежде всего нас будут интересовать предубеждения, связанные с самим научным процессом: пристрастие к получению наглядных либо ярких результатов, подтверждение излюбленной теории или опровержение аргумента, выдвигаемого конкурентом. Любого из них может быть достаточно, чтобы спровоцировать неосознаваемое манипулирование данными, а в некоторых случаях — полное исчезновение неудовлетворительных результатов.

Ирония налицо. Как мы обсуждали, наука олицетворяет наше максимальное приближение к объективности: это процесс, способный преодолеть индивидуальные предубеждения благодаря тому, что чья угодно работа подвергается рецензированию и тщательной проверке. Однако, слишком сосредотачиваясь на этом идеальном представлении о науке как о непогрешимом, беспристрастном методе, мы забываем, что на практике предубеждения преследуют нас на каждом этапе процесса: при чтении предыдущих трудов, продумывании исследования, получении данных, анализировании результатов и принятии решения, публиковать ли их[9]. Тенденция не замечать эти предубеждения превращает научную литературу, которая должна представлять собой точное изложение всего добытого человечеством знания, в крайне субъективный сплав правды и самообмана[10].

Мы начнем эту главу с обсуждения такого предубеждения, которое затрагивает всю научную литературу в целом. Затем присмотримся к тому, как предубеждения влияют на результаты индивидуальных исследований. Для этого нам потребуется совершить небольшой экскурс в статистические методы, чтобы увидеть, как они используются, неверно применяются и неправильно понимаются учеными, анализирующими свои данные. И наконец, мы рассмотрим множество сил, как внутренних, так и внешних, отталкивающих ученых от истины.

Есть такой старый как мир философский вопрос: “Почему существует нечто, а не ничто?” Мы вправе сформулировать похожий вопрос о научном процессе: почему в исследованиях всегда находят что-то, а не ничего? Когда читаешь в газете научный раздел, простительно подумать, что прогнозы ученых постоянно оправдываются, а гипотезы вечно подтверждаются, тогда как исследований, в которых ничего интересного не находят, кот наплакал. Оно и понятно: в газетах ведь должны преподноситься “новости”, а не “летопись абсолютно обо

всем, что произошло". А вот научная литература *должна быть* летописью абсолютно всего, что в науке произошло, — однако демонстрирует то же пристрастие к новым и захватывающим историям. Если просмотреть научные журналы, то обнаружатся бессчетные положительные результаты (когда предсказания ученых оправдываются или обнаруживается что-то новое), но крайне мало результатов отрицательных* (когда исследователи остаются с пустыми руками). Буквально через секунду мы вплотную займемся техническим, статистическим определением "положительных" результатов в противоположность "отрицательным". А пока нам достаточно знать, что ученые обычно стремятся к первым и оказываются разочарованы, если получают вторые.

Специалист по метанауке Дэниел Фанелли в исследовании 2010 года определил, насколько же научная литература позитивна: он перебрал почти две с половиной тысячи статей по всем научным дисциплинам и подсчитал, в скольких из них сообщалось о положительных результатах проверки первой предложенной гипотезы. В различных областях науки уровень позитивности оказался разным. Самым низким (но при этом все равно высоким) — 70,2 % — характеризовались космические исследования, а самым высоким — и вряд ли вас это удивит — психология/психиатрия, где положительные результаты содержались в 91,5 % публикаций[11]. Увязать столь поразительную успешность с уровнем воспроизводимости в психологии, мягко говоря, довольно сложно[12].

Возможно, вас удивляет, почему это нам не следует ожидать высокого уровня успеха в научных исследованиях. В конце концов, ученые обладают фундаментальными знаниями в своей области, и гипотезы обычно являются обоснованными предположениями, а не случайными тычками пальцем в небо. Но, если только ученые не настоящие провидцы, трудно ожидать столь высокого уровня позитивности, какой

* В англоязычной литературе их также называют "нулевыми".

получился у Фанелли. Где же все тупики, те великие идеи, что не выдержали испытания? Где все пробы и ошибки? Где, если уж на то пошло, все ложноотрицательные результаты, когда в ходе исследования не удается получить ожидаемый результат просто по невезению, несмотря на то что гипотеза верна? Иными словами, доля положительных результатов в научной литературе не просто высока, а нереалистично высока[13].

Есть очевидная, но огорчительная причина такой неуемной позитивности: ученые решают, публиковать ли исследование, *на основании его результатов*. В идеальном мире значение имела бы лишь методология исследования: если все согласны в том, что это хорошая проверка сформулированной гипотезы и работа хорошо продуманная, оно публикуется. Это было бы истинным воплощением мертоновской нормы бескорыстности, когда ученым полагается заботиться не о конкретных результатах (сама идея, что у ученых есть свои “излюбленные теории”, — посягательство на эту норму), а только лишь о строгости, с коей проводится изучение.

От реальности это, однако, далеко. Результаты, подтверждающие теорию, оформляются и с помпой подаются в журналы, тогда как разочаровывающие “провалы” (как часто воспринимаются отрицательные результаты) ученые без лишнего шума бросают и переходят к следующему проекту. И это касается не только самих исследователей: редакторы журналов и рецензенты тоже решают, принимать ли статьи к публикации, на основании того, насколько интересно выглядят открытия, и не обязательно обращают внимание на то, сколь дотошны были исследователи в своей работе. И круг замыкается: зачем исследователям утруждать себя подачей в журнал статьи с отрицательными результатами, если ее шансы быть принятой к публикации ничтожны?

Это называют *публикационным смещением*. Оно также известно под старым названием “проблема картотечного ящика”: поскольку в таком ящике ученым предписывалось хранить все свои отрицательные результаты — скрытыми

от лишних глаз[14]. Подумайте об этом в контексте фразы "История пишется победителями", только применительно к научным результатам, либо же так: "Если у вас нет для публикации никаких положительных результатов, не публикуйте ничего вообще".

Чтобы понять, чем публикационное смещение оборачивается на практике, нам нужно подробнее обсудить, как ученые решают, что есть результаты "положительные", а что "отрицательные". А значит, как данные анализируются и интерпретируются. Мы возвращаемся к идее из предыдущей главы, когда мы говорили о сфабрикованных результатах: *в числовых данных всегда есть шум*. Всякому измерению и всякой выборке сопутствует некоторая случайная статистическая флуктуация — ошибка измерения и ошибка выборки. Ее не только трудно подделать человеку — ее еще и непросто отделить от сигнала, который ищут ученые. Зашумленность чисел то и дело подкидывает случайные выбросы и исключения, образуя распределения, которые могут на самом деле быть бессмысленными и обманчивыми. Скажем, когда вы видите очевидное различие по жалобам на боли между группой, принимающей ваше новое лекарство, и контрольной группой, принимающей плацебо, хотя объясняется оно чистой случайностью. Или, казалось бы, вы видите корреляцию между двумя измерениями, которая возникла абсолютно случайно и не появится снова, если исследование повторить. Или вы думаете, что обнаружили энергетический сигнал в ускорителе частиц, а это все случайные флуктуации. Как распознать, где интересующий вас эффект, а где — прихоти случая и ошибок? Для подавляющего большинства ученых ответ кроется в вычислении p-значения.

Откуда берется это p-значение (сокращение от *probability value*, "значение вероятности")? Представьте, например, что мы хотим проверить гипотезу, согласно которой шотландские мужчины выше, чем шотландские женщины. Конечно, мы знаем, что в реальности так и есть: в среднем муж-

чины выше женщин где угодно в мире. Но еще мы знаем, что не всякий мужчина выше всякой женщины; любой из нас вспомнит конкретные случаи, где все наоборот[15]. Предположим, однако, что мы на самом деле не знаем, есть ли в целом разница в росте между мужчинами и женщинами в Шотландии. Тамошнее население составляет лишь пять с половиной миллионов человек, но все равно мы не в силах измерить рост каждого из них, так что для нашего исследования составим случайную выборку более удобоваримого размера. Положим, денег на это исследование у нас маловато, поэтому осилить мы можем только выборку из десяти мужчин и десяти женщин. И вот тут-то в дело и вступает шум. Поскольку рост существенно варьирует от человека к человеку, в итоге у нас случайно — или, если воспользоваться термином, введенным выше, из-за *ошибки выборки* — может получиться группа из необычно высоких женщин и группа из непривычно низких мужчин. Мало того, так как невозможно полностью избавиться от *ошибки измерения*, в нашем распоряжении не будет абсолютно точных значений для роста каждого из участников (вспомните, как мы обсуждали в предыдущей главе, что человек может ссутулиться, рулетка — чуть съехать и так далее).

Скажем, мы обнаружили, что женщины в нашей выборке в среднем на десять сантиметров ниже мужчин[16]. Как нам понять, отражает ли этот результат реальную разницу в популяции (это означало бы, что мы уловили верный сигнал), либо же это только шум (то есть все, что мы видим, случайно)? Нам нужно сравнить две группы в строгом статистическом тесте. Таковых бессчетное множество: *z*-тесты, *t*-тесты, критерии хи-квадрат, отношения правдоподобия и прочие; выбор зависит в том числе от типа данных, с которыми вы работаете. Фактически любой статистический тест сегодня проводится так: вы загружаете свои данные в компьютерную программу и на выходе наряду со многими другими полезными числами получаете соответствующее *p*-значение[17].

Хотя в науке *p*-значение в качестве статистического критерия используется едва ли не чаще всего, общеизвестно, что определение у него весьма мудреное. Недавняя проверка показала, что в целых 89 % учебников по введению в психологию из рассматривавшейся выборки определение дано неверное. Постараюсь не попасть сейчас тоже впросак[18]. *P*-значение — это вероятность, что ваши результаты будут выглядеть так, как они выглядят, или еще лучше, если интересующего вас эффекта на самом деле нет[19]. Важно, что *p*-значение не говорит нам о вероятности того, что результат *верен* (что бы это ни значило), и не сообщает, насколько он *важен*. Оно лишь дает ответ на вопрос: если на самом деле эффект нулевой, каковы были шансы все равно получить те же результаты, что у вас на руках, или демонстрирующие, казалось бы, даже еще больший эффект?[20]

Предположим, в нашем исследовании роста *p*-значение оказалось равным 0,03. Это значит, что если бы в реальности между мужчинами и женщинами в популяции шотландцев не было никакой разницы в росте и мы составляли бы бесконечное число выборок вроде исходных, то лишь в 3 % случаев наблюдали бы разницу в десять сантиметров или больше. В этих 3 % случаев мы бы совершали ошибку, утверждая, что шотландские мужчины в среднем выше, чем шотландские женщины. Иначе говоря, обнаружить разницу в росте, равную или превышающую ту, что получилась для нашей выборки, было бы весьма маловероятно (однако не невозможно), если бы женщины и мужчины в Шотландии на самом деле по росту не отличались.

И поэтому для большинства случаев чем меньше *p*-значение, тем лучше. Но насколько маленьким оно должно быть, чтобы мы уверились в том, что наш результат обусловлен не шумом? Или по-другому: с насколько высокой вероятностью получить ложноположительный результат (когда мы совершаем ошибку, заявляя, что эффект есть, тогда как его на деле нет) должны мы мириться?[21] Чтобы помочь ученым

принимать решения, основоположник статистики Рональд Фишер в 1920-х годах предложил установить пороговое значение, выше которого соответствующий результат будет рассматриваться как отрицательный (ибо слишком уж походит на то, что мы наблюдали бы, если бы ничего на самом деле не происходило), а ниже — как "статистически значимый".

Этот термин породил немало путаницы. Для современного уха "значимый" звучит так, будто указывает на значительный или важный тем или иным образом эффект. Однако, как мы только что обсудили, это не то, что подразумевается под *p*-значением, каким бы малым оно ни было. *Величина* эффекта (показывающая, например, насколько шотландцы выше своих соотечественниц; в нашем примере величина эффекта равняется десяти сантиметрам) — это не то же самое, что вероятность получить такие же результаты, если на самом деле эффекта нет (то есть если мужчины и женщины в Шотландии в среднем одного роста). Скажем, какое-нибудь лекарство, вполне возможно, очень мало помогает при болезни, но вы в достаточной степени уверены, что его эффект не является ложноположительным результатом, — влияние лекарства небольшое, но при этом статистически значимое. Во времена Фишера люди понимали слово "значимый" несколько иначе: подразумевалось, что результат "обозначил" нечто любопытное в данных, хотя, чем бы то ни было, оно не обязательно заслуживало внимания в смысле "реальности"[22].

Так или иначе, Фишер изначально предложил установить "статистически значимый" порог на уровне 0,05 — это значит, что мы должны мириться не более чем с пятипроцентной вероятностью ложноположительного результата в рамках одного испытания (обратите внимание: это означает, что в нашем исследовании, посвященном росту, с *p*-значением 0,03 результат статистически значимый).

Уровень 0,05 выбран довольно-таки произвольно. Похожим образом на великолепном шотландском сайте *taps-aff.co.uk*, где отслеживается погода по всей стране, любая область с тем-

пературой, превышающей 17 градусов Цельсия (около 63 градусов по Фаренгейту), автоматически объявляется зоной "майки долой"* — имеется в виду, что там достаточно тепло, чтобы мужчины имели полное право разгуливать на улице с голым торсом[23]. Семнадцать градусов — вполне разумный порог, но произвольный: пожалуй, некоторые мужчины и не подумают обнажаться, пока температура не поднимется до 20 градусов, а самые закаленные, возможно, начнут скидывать с себя одежду уже при 15 градусах. Поэтому Фишер позднее заметил, что каким-то исследователям, наверное, захочется установить критерий значимости иначе, в зависимости от предмета изучения[24]. Так, порог в "пять сигм", активно обсуждавшийся физиками ЦЕРН после открытия бозона Хиггса в 2012 году, был просто вычурным способом говорить о крайне низком *p*-значении, использовавшемся для столь важного результата[25]: "пять сигм" соответствуют *p*-значению, примерно равному 0,0000003. Вбухав в строительство Большого адронного коллайдера колоссальные ресурсы, физики *действительно* не хотели обмануться шумом в своих числовых данных, поэтому установили очень высокий порог, который доказательство должно было преодолеть.

Если же оставить в стороне исключения вроде бозона Хиггса, порог 0,05 по общему соглашению, традиции и инерции остается на сегодня самым широко используемым критерием. Он заставляет ученых лихорадочно рыться в своих статистических таблицах, отыскивая *p*-значения ниже 0,05, чтобы иметь право сказать: результаты статистически значимы. Легко позабыть о произвольности этого выбора. Ричард Докинз сетовал на "дискретное мышление" — склонность человека мыслить отдельными, строго определенными категориями, а не беспорядочными, размытыми и неоднозначными, которые и характеризуют наш реальный мир[26]. В качестве примера можно привести споры об абортах, когда внимание часто за-

* В оригинале *taps aff* — на шотландском диалекте английского это *tops off*. Отсюда и название сайта.

остряется на том, в какой момент эмбрион или плод становится "личностью", как будто тут вообще возможно провести четкую границу для принятия решения. Аналогично в эволюционной биологии, собственной области интересов Докинза, пытаться указать точный момент, когда один вид эволюционирует в другой, — дело безнадежное, как бы нам того ни хотелось. То же и с *p*-значением: порог отсечения 0,05 для статистической значимости подталкивает исследователей к тому, чтобы считать результаты, преодолевающие эту границу, так или иначе "правдивыми", а остальные — безнадежными "нулевыми". Однако значение 0,05 — это такая же условность, как и "правило 17 градусов" для оголения торса или, если говорить чуть серьезнее, как общественное заключение, что человек официально становится взрослым точно в определенный день рождения.

Прежде чем пуститься в этот несколько затруднительный (но необходимый) статистический обход, мы познакомились с понятием "публикационное смещение" — тенденцией ученых публиковать только положительные результаты и скрывать отрицательные. Теперь мы знаем, как они обычно принимают решение: "значимые" результаты, преодолевающие сакральное пороговое *p*-значение в 0,05, радостно подаются в журналы, тогда как остальные складываются в "картотечный ящик". Отождествление произвольно заданного Фишером статистического порога и "правдивости" или важности результата пагубно повлияло на научную летопись.

Иногда мы замечаем характерные следы публикационного смещения, если рассматриваем весь сегмент научной литературы в целом. Такой широкий взгляд зачастую принимает форму метаанализа, в ходе которого за счет объединения результатов множества исследований, посвященных какой-то одной теме, можно вычислить общий эффект (иногда его называют, пожалуй, искушая судьбу, "истинным" эффектом).

Это может быть, скажем, общее влияние вакцины на снижение смертности от какой-то болезни или общая связь между изменением климата и урожайностью[27].

Подбирая релевантные исследования, специалисты по метаанализу уделяют большое внимание двум показателям. Первый — *величина эффекта*. Для двух наших примеров это будет выглядеть так: вакцина снижает смертность лишь на несколько случаев в год (маленький эффект) или спасает ежегодно тысячи жизней (значительный эффект)? Изменение климата оказывает небольшое, приемлемое влияние на пищевую культуру или же разрушительное? Мы знаем, что из-за ошибки измерения и ошибки выборки различные исследования могут давать очень разные оценки для величины эффекта, поэтому неблагоразумно было бы полагаться на оценку, полученную лишь в одной работе. Поскольку обычно по какому-либо вопросу лучше раздобыть больше доказательств (и поскольку случайные флуктуации, вызываемые ошибками, для разных выборок должны нивелировать друг друга), общая величина эффекта, вычисляемая в ходе метаанализа, считается, как правило, более надежной, чем оценки из отдельных исследований.

При метаанализе общий эффект вычисляется не просто усреднением всех значений для величины эффекта, полученных во включенных туда исследованиях. Учитывается еще и второй показатель, которому специалисты по метаанализу уделяют особое внимание: *размер выборки*. При прочих равных условиях ожидается, что крупные исследования, раз они содержат больше данных, ближе подбираются к “истинному” эффекту (среднему для всей популяции). Иными словами, наилучшая оценка истинного эффекта, полученная в масштабных работах, будет, как правило, точнее, чем оценка из работ небольших[28]. Для нашего исследования жителей Шотландии с выборкой из десяти мужчин и десяти женщин легко представить себе, как мы случайно могли сделать нерепрезентативную выборку из нетипично низких мужчин

или нетипично высоких женщин — и в итоге прийти к неверному выводу. Но вообразите, что мы сделали выборку из тысячи мужчин и тысячи женщин — тогда риск по случайности отобрать тысячу человек с необычной характеристикой, очевидно, гораздо меньше, чем при отборе лишь десяти. Это общее соображение справедливо для большинства ситуаций: в небольших исследованиях, которые представляют собой более ограниченные срезы реальности и сильнее подвержены влиянию ошибки выборки, вариабельность будет выше, то есть истинный эффект будет либо переоценен, либо недооценен больше. Вот почему при метаанализе больший вес придается величине эффекта, получившейся в масштабных исследованиях, ведь обычно они точнее[29].

Что же касается публикационного смещения, нас интересует, как величина эффекта и размер выборки друг с другом соотносятся. Если в одной системе координат по осям отложить два этих параметра, так чтобы каждая точка соответствовала одному исследованию, логично будет ожидать увидеть что-то вроде диаграммы, изображенной на рисунке 2А. (Обратите внимание: это идеализированная версия метаанализа, реальные наборы данных почти никогда не выглядят столь четко.) Глядя на эту “воронкообразную диаграмму” (названную так по очевидной, надеюсь, причине), вы можете заметить, насколько сильно разбросаны точки, соответствующие небольшим исследованиям, то есть располагающиеся ближе к горизонтальной оси. По мере того как мы поднимаемся по вертикали, точки, соответствующие уже более масштабным исследованиям, начинают собираться вокруг среднего эффекта, иллюстрируя принцип, который мы только что обсудили: более масштабные исследования точнее. Разброс вдоль горизонтальной оси показывает, почему принимать как данность эффект, оцененный в любом отдельном исследовании, — плохая идея: несмотря на то что в данном примере есть реальный эффект, отдельные исследования искажают его “истинную” величину в меньшую или большую сторону в раз-

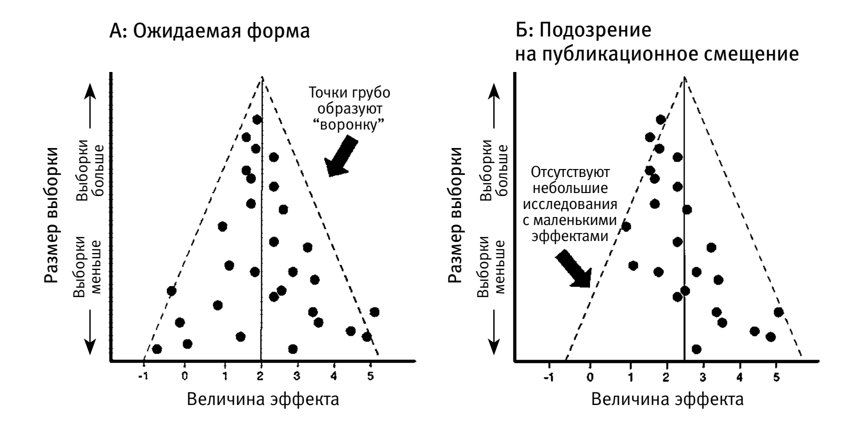


Рис. 2. Два разных варианта воронкообразной диаграммы, полученные в ходе воображаемых метаанализов. На диаграмме А распределение тридцати точек примерно такое, какого следовало бы ожидать, если бы результаты каждого когда-либо проводившегося исследования на данную тему были опубликованы. На диаграмме Б шесть точек из левой нижней области (исследования с маленькими выборками и небольшими эффектами) пропали — такая картина может быть признаком публикационного смещения. Вертикальная линия посередине каждой из двух диаграмм показывает общую величину эффекта, вычисленную в ходе соответствующего метаанализа. На диаграмме Б она сдвинута вправо, а значит, в ходе этого метаанализа эффект получен больше, чем он должен быть.

личной степени (тогда как самые масштабные исследования демонстрируют несравненную точность). В любом случае все кажется закономерным: форма перевернутой воронки — это именно то, чего и следовало бы ожидать, если бы все исследования сходились к реальному эффекту.

Как при археологических раскопках, когда *отсутствие* определенных предметов сообщает любопытную информацию об изучаемых исторических персонажах (например, отсутствие оружия указывает на то, что они были скорее гражданскими лицами, чем военными), мы можем многое узнать, проанализировав, чего в метаанализе не наблюдается. Что, если наша диаграмма больше походит на рисунок 2Б, ко-

гда утерян участок ожидаемой формы? Точки, соответствующие исследованиям с маленькими выборками и небольшими эффектами, которые мы бы ожидали увидеть в левой нижней части воронки, пропали. Рассуждая как археолог, специалист по метаанализу вправе заключить, что такие исследования проводились, но их результаты авторы, вместо того чтобы опубликовать, погребли в "картотечном ящике". Почему? Вероятное объяснение: в этих исследованиях с маленькими выборками и небольшими эффектами *p*-значения получились выше 0,05 — и результаты сочли бесполезными, отрицательными.

Наверное, ученые, проводившие исследования, рассуждали примерно так: "Что ж, исследование небольшое, поэтому обнаруженный маленький эффект объясняется, пожалуй, просто зашумленными данными. Если подумать, глупо было вообще *надеяться* обнаружить здесь какой-то эффект! Нет никакого смысла публиковать эти результаты". Важно, однако, что они не занимались бы подобной рационализацией задним числом, если бы то же самое исследование при небольшой выборке — с его потенциально зашумленными данными — показало значительный эффект, а охотно отослали бы свои положительные результаты в журнал. В этих двойных стандартах, в основе которых лежит неистребимая в человеке склонность к подтверждению своей точки зрения (интерпретировать данные так, чтобы подтверждались собственные априорные убеждения и ожидания), и коренится публикационное смещение.

При сравнении общих выводов из метаанализов, соответствующих рисункам 2Б и 2А, видно, как публикационное смещение искажает научную литературу. Если из воронкообразной формы удалены исследования с маленькими эффектами, то общий эффект, который получится в метаанализе, будет по определению больше оправданного. Мы получаем завышенную оценку важности эффекта и можем прийти к неверному заключению, поверив, что нечто существует, хотя

на деле это не так. Не публикуя отрицательные или неоднозначные результаты исследований, ученые нацепляют шоры на любого читателя научной литературы.

Одна из самых поразительных недавних воронкообразных диаграмм была получена в метаанализе психологом Дэвидом Шэнксом и его коллегами[30]. Они изучали еще одну вариацию прайминга — "романтический прайминг". Согласно этой идее, после того как мужчинам показали фотографию привлекательной женщины, они рвутся принимать на себя бóльшие риски и больше тратить на потребительские товары (занимаясь "показным потреблением" для привлечения партнеров). Пятнадцать опубликованных на эту тему статей, описывающие в общей сложности сорок три отдельных эксперимента, вроде бы подтверждали такую гипотезу. Тем не менее, когда в ходе метаанализа по точкам, соответствующим этим исследованиям, была построена диаграмма, от воронки оказался откушен огромный кусок: веское доказательство, что многие из работ, в которых эффект не проявился, опубликования не удостоились. И действительно, когда Шэнкс с коллегами попытался воспроизвести эффект романтического прайминга в собственных масштабных экспериментах, тот вообще никак не проявился — значения для величины эффекта во всех повторах стремились к нулю.

Не менее ярко публикационное смещение выражено в медицине. Так, в анализе 2007 года обнаружилось, что в более чем 90 % статей, описывающих эффективность прогностических тестов при раке, сообщалось о положительных результатах. В реальности мы до сих пор не особенно хорошо предсказываем, у кого будет рак, а стало быть, что-то с этой литературой не в порядке[31]. В другом исследовании, когда изучалось сорок девять метаанализов, посвященных потенциальным маркерам заболеваний сердечно-сосудистой системы (например, в крови тех, кто в группе риска по инфаркту, определенных белков может присутствовать больше), было показано, что в целых тридцати шести из них есть следы пристрастия

к положительным результатам[32]. Иными словами, опубликованные статьи явно сильно раздували кажущуюся полезность этих биомаркеров.

То же касается и средств лечения. Врачам, прописывающим лекарства своим пациентам, нужно соизмерять пользу и вред, например, решая, стоит ли человеку принимать антидепрессанты, несмотря на их распространенные побочные эффекты вроде тошноты и бессонницы. Если медицинская литература создаст у врачей чересчур оптимистичное представление о пользе лекарства (что, похоже, справедливо для антидепрессантов, которые и правда работают, но оказывают не такой сильный эффект, как изначально предполагалось), их клинические обоснования окажутся искаженными[33].

Совершенно естественно, если прежде вы о публикационном смещении никогда не слышали: это один из самых постыдных секретов науки. Однако в 2014 году при изучении обзоров из топовых медицинских журналов выяснилось, что в 31 % метаанализов даже не проверялось его возможное наличие. (Когда же все тщательно *проверили*, в 19 % из этих метаанализов обнаружилось, что публикационное смещение действительно имело место[34].) Более позднее изучение обзоров, посвященных исследованиям рака, дало результат еще хуже: в 72 % не проводилось проверок на публикационное смещение[35]. Зачастую непонятно, что в точности нужно делать, если в вашем наборе данных для метаанализа обнаружились признаки публикационного смещения (следует ли снизить оценку среднего эффекта, и если да, то насколько?), но вряд ли верное решение заключается в полном игнорировании проблемы[36].

Затруднение с археологическим подходом к публикационному смещению состоит в следующем. Он опирается на предположение, что надо заполнить лакуны на воронкообразной диаграмме — те области, где мы ожидаем увидеть точки, соответствующие небольшим исследованиям с маленькими эффектами, но воронкообразные диаграммы

могут иметь странные формы по причинам, не связанным с публикационным смещением, особенно если между разношерстными исследованиями, отобранными для метаанализа, много различий[37]. Есть немало случаев, когда публикационное смещение менее явное, чем в обсуждавшихся выше примерах, а стало быть, его сложнее распознать. Нет ли более надежных способов проверить, не имеет ли место такого рода предвзятость?

Альтернативный подход таков: нужно взять набор исследований, которые, по вашим сведениям, точно были завершены, со спектром результатов от существенно положительных до отрицательных, а затем проверить, сколько из исследований каждого типа оказались опубликованными. Именно это в 2014 году и сделала группа стэнфордских ученых, возглавляемая политологом Энни Франко, в ходе проекта, названного ими “вскрытием картотечного ящика”[38]. Они рассматривали исследования, чьи авторы подавали заявки на участие в правительственной программе, поддерживающей обзорные работы[39]. Используя успешные заявки, относящиеся к периоду с 2002 до 2012 года, в качестве реестра тех работ, что прошли на следующий этап, Франко и ее коллеги проследили судьбу каждого исследования, связываясь, если требовалось, с авторами напрямую. В итоге выяснилось, что в 41 % завершенных исследований нашлись четкие подтверждения исходной гипотезы, в 37 % результаты оказались неоднозначными, а в 22 % — отрицательными. В мире, где значение имеют методы, а не результаты, процентное соотношение среди опубликованных статей было бы сходным. Но ничуть не бывало. Среди опубликованных статей доли тех, что сообщали о положительных, неоднозначных либо же отрицательных результатах, равнялись 53, 38 и 9 % соответственно. Иными словами, вероятность оказаться опубликованными для положительных результатов на 44 % превышала таковую для отрицательных[40].

От соответствующих ученых Франко и ее коллеги узнали, что 65 % исследований с отрицательными результа-

тами вообще даже и не оформлялись в виде статьи, не говоря уже об отправке в журнал. Многие из этих ученых предвидели, что опубликоваться у них нет никаких шансов. “Такова суровая реальность мира публикаций: нулевые эффекты не рассказывают четкой истории”, — сказал один. “Учитывая сильное предпочтение, которое в этой области отдается положительным результатам, дальше я не пошел”, — объяснил другой. Многие просто махнули рукой и переключились на следующие проекты. Один ученый признался: “Я так и не сел за написание статьи из-за того, что было некогда, интерес несколько поугас… да и важных результатов не имелось”[41]. Другими словами, “картотечный ящик” действительно существует[42].

В бизнесе и политике одни из самых отъявленных злодеев — подпевалы. Бесчисленные книги учат начинающих управленцев и руководителей ни в коем случае не окружать себя людьми, которые просто будут поддакивать всем решениям, даже плохим. Уинстон Черчилль писал, что искушение говорить высокому начальству только те вещи, которые оно больше всего хочет услышать, — это одно из самых распространенных объяснений ошибочной политики, а значит, лидеру, от чьих решений зависят судьбоносные события, перспективы обычно видятся куда более оптимистичными, чем допускают строгие факты[43]. В науке таких подпевал делает из выходящих в журналах статей публикационное смещение: мы видим все положительные, но не можем узреть строгие отрицательные результаты. Принятие решений на основании такой частичной информации — прямой путь к катастрофе.

Немаловажно, что против публикационного смещения есть и моральный довод. Если вы провели исследование на людях, особенно если участники принимали лекарство или проходили экспериментальное лечение, то вы, можно сказать, обязаны ради них опубликовать результаты. Иначе получится, что все неудобства, через которые им пришлось пройти (включая, как иногда бывает, болезненные процедуры или побоч-

ные эффекты), были напрасны. Похожий аргумент применим к исследованиям, которые вы проводили на чьи-то деньги.

Итак, с какой стороны ни посмотри (с научной, практической или этической), публикационное смещение — большая проблема. К сожалению, это далеко не единственная проблема, порождаемая неистребимым, глубоким пристрастием науки к положительным результатам.

Для ученого амбициозного, ориентированного на карьерный рост публикационное смещение имеет существенное негативное последствие: пряча отрицательные результаты в "картотечный ящик", он лишается этой жизненно важной публикации и столь отрадной дополнительной строчки в резюме. Чтобы избежать потерь и спасти свои результаты, есть другой выход — манипулирование данными. Речь не о явной фальсификации, выдумывании данных, как в предыдущей главе, а скорее о некоем неосознаваемом (или полуосознаваемом) искажении данных — по выражению Гулда, "слабо ощущаемом жульничестве", — в которое ученые могут впасть совершенно нечаянно, без задней мысли. Действительно, манипулирование данными пугает не только тем, что запруживает научную литературу неверными выводами, но еще и тем, что так много ученых либо совершают его абсолютно непреднамеренно, либо, если понимают, что манипулируют данными, не задумываются, почему это плохо.

Полагаю, вы слышали об идее, согласно которой, взяв в столовой большую тарелку, вы съедите больше, чем если бы взяли маленькую. А значит, вы косвенно знаете профессора Брайана Вонсинка. Руководитель лаборатории по изучению еды и брендов в Корнеллском университете, Вонсинк долгое время был, несомненно, самым большим мировым авторитетом в области психологии питания. На его счету популярные книги, двухгодичное управление Центром политики и продвижения в области питания при министерстве сельского хо-

зяйства США во времена президентства Джорджа Буша — младшего и сотни публикаций, причем многие цитировались в контексте инициативы "Обеды с умом", осуществлявшейся в американских школах в эпоху Обамы[44]. Вонсинк даже выиграл Шнобелевскую премию — шутливый аналог Нобелевской, отмечающий исследования, которые "заставляют людей сначала засмеяться, а потом задуматься". Было это в 2007 году, его наградили за исследование, в котором он склонял людей потреблять куда больше супа, чем они собирались, с помощью суповой миски, смонтированной так, что она непрерывно вновь себя наполняла[45]. Способность Вонсинка представлять забавные запоминающиеся результаты по психологии питания казалась безграничной, сродни бездонной миске. Помимо своих результатов на тему размера порций он обнаружил, что если вы пошли в магазин за продуктами голодным, то купите больше калорий, что взгляд персонажей, изображенных на упаковках со сладкими хлопьями, часто направлен вниз, словно чтобы установить зрительный контакт с маленькими детьми, стоящими перед полкой в супермаркете, и что приклеивание на яблоко стикера с Элмо из "Улицы Сезам" заставляет детей чаще выбирать яблоко вместо печенья[46].

И лишь в конце 2016 года все рассыпалось в прах. Вонсинк написал сообщение в блоге о том, как он велел одной своей студентке проанализировать набор данных, собранных им в нью-йоркской пиццерии[47]. Исходная гипотеза, сказал он, "не подтвердилась", но вместо того, чтобы опубликовать отрицательные результаты или вообще все положить в стол, он уверил ее, что "из этих данных точно можно что-то вытянуть". Она подчинилась и "каждый день… возвращалась со странными новыми результатами… и ежедневно… придумывала новый способ проанализировать данные заново". Открыто призвав прочесывать свои наборы данных, выискивая все "значимое", Вонсинк нечаянно обнажил серьезный недостаток того способа, каким он и, увы, многие тысячи других ученых проводили исследования.

Этот недостаток прозвали “*p*-хакингом”*[48]. Поскольку критерий “$p < 0{,}05$” так важен для того, чтобы статья была опубликована (в конце концов, он якобы сигнализирует о “реальности” эффекта), ученые, чьи исследования дают неоднозначные или разочаровывающие результаты, регулярно прибегают к ухищрениям, которые самую малость сдвигают их *p*-значения ниже этого заветного порога. Такой *p*-хакинг бывает двух основных типов. При первом ученые, придерживающиеся определенной гипотезы, прогоняют, и перепрогоняют, и переперепрогоняют свой анализ результатов эксперимента каждый раз немножко другим способом, пока им в конце концов не выпадает удача в виде *p*-значения меньше 0,05. Они могут вносить изменения в свой импровизированный анализ по-всякому: выбрасывать определенные точки, перевычислять значения внутри отдельно взятых подгрупп (например, проверять, нет ли эффекта, только для мужчин, потом для женщин), перебирать различные виды статистических тестов или продолжать набирать новые данные, совершенно не собираясь останавливаться, до тех пор пока *где-то* порог по значимости не окажется перейден[49]. При другом варианте *p*-хакинга берется имеющийся набор данных, прогоняется через кучу произвольных статистических тестов без какой-либо определенной гипотезы в голове, а после всем просто докладывается о каких угодно эффектах, которым случилось получить *p*-значения ниже 0,05. Затем ученый может заявить (не исключено, что часто убеждая даже самого себя), что изначально и ждал этих результатов[50]. Этот последний вид *p*-хакинга известен как “харкинг” (*HARKing*, *Hypothesising After the Results are Known*, то есть “формулирование гипотезы уже после получения результатов”). Его наглядно демонстрирует часто повторяемая аналогия про “техасского стрелка”, который выхватил револьвер, наугад изрешетил стену амбара, а потом, нарисовав мишень вокруг нескольких пулевых отверстий, по случай-

* Хакинг в смысле “отслеживание”, “выуживание”. На английском *p-hacking*.

ности оказавшихся поблизости друг от друга, бахвалится, что, мол, туда и целился[51].

Оба вида *p*-хакинга есть проявления одной и той же ошибки, и по иронии — именно той, ради избежания которой и были изобретены *p*-значения: ошибки извлечения выгоды из абсолютной случайности. Несмотря на то что *p*-значения призваны помочь нам отделить сигнал от шума, эта их способность сбоит, когда вы вычисляете большое их количество для одного и того же исследуемого вопроса. Всякий раз, как вы подлаживаетесь под свои данные или под статистический тест, вы даете себе еще один бросок костей, увеличивая свои шансы на то, что выпадет случайная флуктуация и вы объявите ее "реальной". Как мы видели, пороговое *p*-значение в 0,05 означает, что если наша гипотеза неверна (например, новое лекарство на самом деле не работает), то в 5 % случаев мы будем получать ложноположительный результат, то есть объявлять, что эксперимент увенчался успехом, когда это не так. Но эти 5 % — для *единичного испытания*. Нехитрые математические вычисления показывают, что в мире, где наша гипотеза неверна, увеличение числа проводимых статистических тестов повышает наши шансы получить ложноположительный результат[52]. Если мы проводим пять (не связанных друг с другом) тестов, то вероятность получить хотя бы один ложноположительный результат — 23 %, если тестов двадцать, то 64 %. Следовательно, при многократных тестах мы выходим далеко за допустимый уровень в 5 %, предложенный Фишером. Вместо того чтобы помогать нам вычленить сигнал, *p*-значения отбрасывают нас в море шума.

Нельзя сказать, что эта идея прямо-таки интуитивно понятна, поэтому приведу пример, используя обожаемую статистиками аналогию. Пусть у меня есть мешок с монетами, но я беспокоюсь, что все они утяжелены так, чтобы орлы выпадали чаще, чем решки. Я вынимаю одну монету, подбрасываю пять раз, и выпадают одни орлы. Это послужило бы как минимум умеренно убедительным доказательством того, что про-

исходит нечто странное. А теперь допустим, что на этой первой монете орлов выпало только три, а решек две — не сверхубедительное доказательство моей гипотезы. Однако вместо того, чтобы отбросить гипотезу, я мог бы отреагировать так: начать проверять другие монеты, выуживая их одну за другой из мешка и подбрасывая, до тех пор пока на одной из них наконец-то не выпадет пять раз подряд орел. Надеюсь, вы согласитесь, что эта ситуация куда менее убедительна, чем та, когда первая же проверенная мной монета дала сплошных орлов. Но что, если я прикрою свои многократные испытания некой историей? "На самом деле, — мог бы я сказать, — каждый раз, вытаскивая новую монету из мешка, я проверял новый вариант своей гипотезы. Во второй раз я проверял, не утяжелены ли монеты таким образом, что орлы выпадают только при подбрасывании левой рукой. В третий раз я проверял, не происходит ли это, только когда комнатная температура превышает двадцать градусов. В четвертый раз…" — ну, вы поняли. Я мог бы убедить даже самого себя, будто и вправду проверял все эти новые интересные гипотезы. Но по большому счету я просто добавлял себе попыток в одном и том же тесте, тем самым повышая свои шансы получить ложноположительный результат. И как только я нашел одну монету, которая дала пять заветных орлов подряд, я могу испытать искушение просто опубликовать данные лишь по ней одной.

Та же логика превращает потрясающие, казалось бы, совпадения вроде такого: "Я вдруг подумал о человеке, с которым уже много месяцев не общался, как в ту же секунду он присылает мне сообщение!" — в менее впечатляющие, как только вы примете во внимание, сколько тысяч, а то и миллионов людей по всему миру спонтанно думают о ком-то и *не* получают неожиданных сообщений. В конце концов, шанс один на миллион выпадает порядочно часто, если ваша популяция состоит из нескольких миллионов людей. Увеличьте число возможностей для выпадения случайного результата — и можете биться об заклад, что рано или поздно он таки выпадет; вытянутые

из громадных множеств особые значения не доказывают, что это не просто случайность. Несомненно, даже если вы честно провели только один тест и получили *p*-значение меньше 0,05, вы все равно можете случайно обмануться. Но этот риск существенно ниже, чем в случае *p*-хакинга, когда проведение многочисленных тестов увеличивает риск, что какой-то из них введет в заблуждение.

Это важнейшая идея, которую так много ученых, похоже, не улавливают: даже когда ничего не происходит, вы все равно регулярно получаете "значимые" *p*-значения, особенно если проводите уйму статистических тестов[53]. Как с "экстрасенсами": они делают тысячи предсказаний о событиях на следующий год, а затем в конце года подчеркивают лишь те, насчет которых оказались правы, тем самым придавая всему такой вид, словно они обладают магическими способностями предсказывать будущее[54]. Бросьте статистические кости достаточное число раз — и что-то вроде бы знаменательное проявится, пусть даже это просто нелепый выброс в ваших данных[55], а еще никому не рассказывайте обо всех тех разах, когда открытие не было значимым, — и вот вам прекрасный рецепт, как убедить людей в реальности своего результата, хотя основан он не более чем на шуме.

Это возвращает нас к Брайану Вонсинку, его студентке и данным по пиццерии. Хотя при всех ежедневных попытках студентки провести анализ данных вычислялось великое множество *p*-значений, лишь несколько из них попали в итоге в опубликованную литературу. В соответствии с неписаными правилами мира научных публикаций, эти значения были преимущественно меньше 0,05. Мы, читатели научной литературы, понятия не имеем, сколько тестов было проведено. Поскольку так много результатов оказалось сокрыто от посторонних глаз, это сродни публикационному смещению внутри одного исследования. Если бы мы могли увидеть весь процесс с отрицательными результатами и всем прочим, он выглядел бы как классический случай с техасским стрел-

ком. В своем злосчастном сообщении в блоге Вонсинк ненароком дал понять, что процесс включал в себя рисование мишени вокруг результатов, которым случилось — по стечению обстоятельств — оказаться статистически значимыми. Для тех, кто понимает, как работает статистика, мишень Вонсинка красовалась не на стене амбара: она пылала на его собственной ноге.

После того как Вонсинк опубликовал свое сообщение в блоге, некоторые скептически настроенные читатели начали разбираться с числами в его статьях[56]. Выяснилось, что *p*-хакинг был лишь одним из массы статистических проколов. В четырех статьях Вонсинка, основанных на данных по пиццерии, команда скептиков нашла не менее ста пятидесяти ошибок — целый набор расхождений между четырьмя этими исследованиями (а иногда числа противоречили друг другу внутри одной и той же статьи)[57]. Все стало еще хуже, когда скептики начали проверять другие работы Вонсинка. Повторный анализ данных из одной статьи о кулинарных книгах показал, что почти каждое число переврано[58]. В публикации о наклеивании на яблоки стикеров с Элмо Вонсинк неправильно подписал график и неверно описал методологию исследования[59]. Вскоре начался отзыв статей: на момент написания этих строк восемнадцать публикаций Вонсинка были изъяты из научной литературы, и есть подозрения, что их станет еще больше[60]. Чуть меньше чем через два года после того пресловутого сообщения в блоге Вонсинк подал заявление об увольнении из Корнеллского университета[61].

Возмутительные ошибки и просчеты Вонсинка отвлекали внимание от детали его авантюры, распространенной шире всего, — от *p*-хакинга. Посреди скандала, разразившегося в средствах массовой информации, когда начался отзыв статей, журналист из *BuzzFeed News* обнародовал электронное письмо Вонсинка одному из соавторов, отправленное во время написания ими треклятой статьи про наклейки с Элмо, — крайне откровенное и обличительное. В нем Вонсинк беспокоился, что, “хотя наклейки повышают вероят-

ность выбора яблок на 71 %, по какой-то причине *p*-значение равно 0,06. Мне кажется, оно должно быть меньше. Не хотите ли взглянуть и сказать, что вы об этом думаете? Если вы можете взять данные и их потребуется немножко поткрутить [*sic*], желательно было бы сделать это значение меньше 0,05"[62].

Это редкий случай, когда ученый открыто призывает своих коллег к *p*-хакингу. Но поразительным он был только из-за прямоты. Подозреваю, многие ученые, когда история Вонсинка была предана огласке, нервно заерзали в своих креслах, понимая, что он просто оказался на краю спектра, которому принадлежат и они сами. Возможно, они не так небрежны, как Вонсинк, и в своих электронных письмах (или при личном общении, когда не сохраняется письменный след) деликатнее просят проанализировать данные заново[63]. Но когда желание "сделать это значение меньше 0,05" сильно — а так и есть, ведь журналы явно отдают предпочтение интересным, ярким положительным результатам, — *p*-хакинг почти неизбежен.

Из-за нечаянного признания все закончилось для Вонсинка плохо. Однако, когда очередной известный ученый откровенно признал, что в прошлом невольно прибегнул к *p*-хакингу своих данных, реакция научного сообщества была обнадеживающей. Вспомним про обсуждавшуюся во второй главе невероятную успешность идеи "поз силы", основанной на статье 2010 года, результаты которой не воспроизвелись. Несмотря на то что имя Эми Кадди стало прочно ассоциироваться с этим понятием, ведущим автором статьи была вообще-то не она. Это была Дана Карни из Калифорнийского университета в Беркли, и в 2016 году она выпустила заявление о своих переменившихся взглядах на позы силы. В течение прошедших до того момента лет она корректировала свои представления — как она сама выразилась, вплоть до позиции "я не верю, что «позы силы» действительно работают". Далее она перечислила некоторые факты об исходном эксперименте (с "крошечным" размером выборки в сорок два

человека и “едва уловимой” величиной эффекта), в совокупности складывающиеся в ясную как день историю *p*-хакинга:

- Они набирали участников “подолгу и попутно проверяли эффект” (то есть продолжали увеличивать выборку, пока не получили значимый результат).
- Нескольких участников исключили по причинам, кажущимся произвольными.
- Некоторые выпадающие точки были удалены из рассмотрения, а другие — оставлены.
- Использовались различные показатели и различные статистические тесты, но сообщалось только о тех, которые давали самые маленькие *p*-значения.
- Об оценке участниками собственной силы задавалось множество вопросов, однако сообщалось только о тех, что показали эффект[64].

По словам Карни, “тогда это не казалось *p*-хакингом”, хотя это определенно был он. Обрушился ли на нее град страшных ругательств, потеряла ли она работу после такого признания? Нет. На самом деле реакция оказалась в точности обратной. Поиск в твиттере, часто поносимом как рай для онлайн-травли, показывает, что другие исследователи (справедливо) назвали заявление Карни “смелым”, “впечатляющим”, “достойным восхищения”, “шагом вперед”, примером того, “как нужно действовать, если провалилась попытка воспроизвести какую-то работу”, “выдающейся демонстрацией научной прямоты”. Один специалист по нейронауке назвал Карни “интеллектуальным/академическим героем”[65]. Мне не удалось найти ни одной негативной реакции — кроме той, что дала сама Эми Кадди, воспользовавшись возможностью отмежеваться от *p*-хакинга: “Я… не могу оспаривать воспоминания первого автора [то есть Карни] о том, как данные собирались и анализировались, поскольку она руководила обоими процессами”[66].

Насколько распространены подобные аналитические предубеждения, что пронизывали работу Вонсинка и подорвали исследование поз силы? В 2012 году в ходе опроса свыше двух тысяч психологов выяснялось, не вовлекались ли они в какие-либо практики *p*-хакинга[67]. Приходилось ли им когда-нибудь собирать данные по нескольким разным исходам, но не сообщать обо всех? Около 65 % сказали, что да. Исключали ли они определенные точки из анализа после того, как уже взглянули на результат? В этом сознались 40 %. И примерно 57 % сказали, что принимали решение подсобрать еще данных, после того как провели анализ — и, видимо, сочли его неудовлетворительным.

Опросы в других сферах принесли сходные удручающие открытия. Опрос статистиков в области биомедицины 2018 года выявил, что к 30 % респондентов их клиенты-ученые обращались с просьбой "интерпретировать статистические данные исходя из ожиданий, а не фактических результатов", при этом 55 % участников опроса их клиенты просили "подчеркнуть только значимые результаты, а о незначимых умолчать"[68]. В другом опросе 32 % экономистов признались, что "представляли эмпирические результаты избирательно, так чтобы они подтверждали их аргументы", а 37 % сказали, что "прерывали статистический анализ, когда получали [свой] результат" — то есть в тот момент, когда *p*-значению случалось опуститься ниже 0,05, хотя это могло произойти просто по стечению обстоятельств[69]. Если собрать вместе все *p*-значения из опубликованных работ и построить диаграмму, то мы увидим странный резкий скачок *чуть пониже* 0,05: значений 0,04, 0,045, 0,049 и так далее будет куда больше, чем можно было бы ожидать по воле случая. Аргумент несокрушительный, однако это красный флаг, сигнализирующий о возможном *p*-хакинге: ученые, похоже, подкручивают свои исследования как раз настолько, чтобы результаты поднырнули под линию в 0,05, а затем отсылают их в журналы[70].

Нужно помнить, что *p*-хакинг может ощущаться учеными так, словно они делают свои результаты, кажущиеся им правдивыми, некоторым образом более четкими или реалистичными. И это снова ошибка подтверждения. Вот тот участник? Я *точно* видел, как он глазел в окно, вместо того чтобы сосредоточиться на выданном ему тесте по психологии. Вот та чашка Петри? На ней *определенно* виднелось пятнышко грязи, так что она, вероятно, контаминирована, лучше выкинуть соответствующие результаты из набора данных. *Нет сомнений:* логичнее провести статистический тест X вместо теста Y (и гляньте-ка, получается, что статистический тест X дает положительные результаты!). Думаю, идею вы уловили. Как мы обсуждали в предыдущей главе, когда рассматривали мотивы нечистых на руку ученых, если вы еще до проверки своей гипотезы верите, что она верна, вам может показаться в высшей степени уместным подтолкнуть любые неоднозначные результаты в нужном направлении. И если настоящий мошенник знает, что поступает неэтично, обычные исследователи, занимающиеся *p*-хакингом, часто этого не сознают.

Не бывает так, чтобы набор данных можно было проанализировать лишь одним-единственным способом. Исключаете ли вы из рассмотрения выпадающие точки, поскольку полагаете, что они делают вашу выборку менее репрезентативной, хуже отражающей популяцию в целом, или же оставляете их? Разделяете ли вы выборку на разные группы по возрасту либо по какому-то другому критерию? Объединяете ли вы наблюдения первой и второй недели и сравниваете их с наблюдениями за третью и четвертую или рассматриваете каждую неделю по отдельности, а может, группируете их как-то иначе? Выбираете ли вы эту конкретную статистическую модель или ту? Сколько в точности “контрольных” переменных вы включаете в рассмотрение? На такого рода вопросы нет однозначных ответов — они зависят от особенностей и смысла исследования и от вашего взгляда на статистику (она все-таки и сама по себе постоянно развивающаяся область). Спроси де-

сять статистиков — и получишь, вероятно, столько же разных ответов. Эксперименты по метанауке, в которых нескольким исследовательским группам поручается проанализировать один и тот же набор данных или придумать с нуля дизайн собственного исследования для проверки одной и той же гипотезы, показали высокую степень вариативности в методах и результатах[71].

Неограниченный выбор предоставляет неограниченные возможности для тех ученых, кто приступает к анализу, не имея в голове четкой идеи искомого. Но, как теперь должно быть понятно, чем больше анализируешь, тем выше шансы получить ложноположительные результаты. Специалисты по работе с данными Таль Яркони и Джейк Вестфолл объясняют это так: "Чем менее строгим... исследователь хочет быть — то есть чем шире спектр закономерностей, которые он хочет «увидеть» в данных, — тем существеннее риск попасть в ситуацию, когда ему привидится закономерность, коей там вовсе и нет"[72].

Дальше — хуже. Пока звучало все так, будто *p*-хакинг всегда осуществляется явным образом — прогоном множества статистических анализов и обнародованием только тех, что дают *p*-значения ниже 0,05. Вне всякого сомнения, такое происходит часто, однако настоящая проблема куда щекотливее. Вот она: даже если вы проводите лишь один анализ, вам все равно надо учитывать все те, что вы *могли* провести. Статистики Эндрю Гелман и Эрик Локен сравнили процесс проведения незапланированного статистического анализа с "садом расходящихся тропок" из рассказа Хорхе Луиса Борхеса с одноименным названием: в каждой точке, где требуется принять аналитическое решение, вы можете выбрать какую угодно из множества имеющихся опций. Любой выбор, как мы видели, приведет к несколько иным результатам[73]. Пока вы не установили очень четкие критерии того, как должен выглядеть результат, подтверждающий вашу гипотезу, пока вы не сказали, что хотите получить "*p*-значение меньше 0,05 при

вот *такой* работе с переменными при вот *этих* конкретных условиях и с *такими* вот контролями", вы будете в итоге принимать за доказательство своей правоты любой из многочисленных возможных результатов. Но откуда вам знать, не является ли тот результат, с которым вы остались, пройдя уникальную комбинацию расходящихся тропок, статистической случайностью? Таким образом, даже вне проб и ошибок классического *p*-хакинга ученые, подходящие к своим данным без тщательно продуманного плана, могут в итоге загнать себя в угол невоспроизводимости.

Почему "невоспроизводимости"? Да потому что, когда ученый достигает каждого разветвления на тропке, он оказывается одурачен данными — делая выбор, который кажется ведущим к $p < 0{,}05$ *в этом наборе данных*, но не обязательно приведет туда же в других наборах данных. Это проблема всех видов *p*-хакинга, будь он явный или нет: они приводят к *чрезмерной аппроксимации* (если использовать технический термин, или же "переподгонке") данных в процессе анализа[74]. Другими словами, анализ, положим, описывает картину в этом конкретном наборе данных хорошо, но сама картина может быть обусловлена шумовыми вывертами или странностями, которые не поддаются обобщению на другие данные или на реальный мир. Это попросту бесполезно, ведь чаще всего нас не интересует внутренняя кухня отдельно взятого набора данных (мы не хотим узнать, "какова связь между приемом нейролептиков и симптомами шизофрении, измеренная в этой конкретной выборке из двухсот трех человек с апреля по май 2019 года в Денвере, штат Колорадо") — мы ищем обобщаемые факты о мире ("какова связь между приемом нейролептиков и симптомами шизофрении у людей в целом?").

На рисунке 3 проиллюстрирована чрезмерная аппроксимация. Как вы видите, у нас есть набор данных: одно измерение количества осадков за каждый месяц на протяжении года. Мы хотим провести через наши данные линию, которая описывала бы, как меняется количество осадков со временем, —

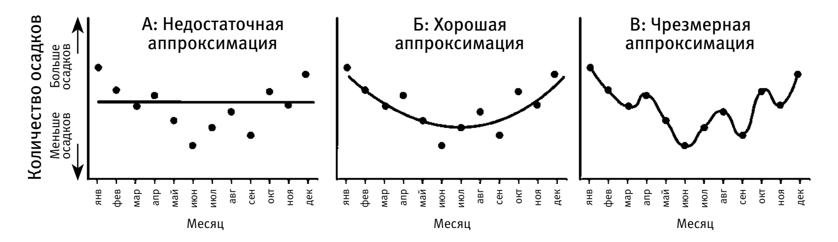


Рис. 3. Проблема переподгонки, проиллюстрированная с помощью моделей для осадков. На графике А модель плохая — это “недостаточная аппроксимация”, поскольку она неудовлетворительно описывает данные. На графике Б модель куда лучше, ведь она описывает выпадение осадков таким образом, который обобщается, вероятно, и на другие годы. На графике В показана “чрезмерная аппроксимация”: хотя модель прекрасно описывает данные именно этого года, маловероятно, что в другие годы значения будут скакать вверх и вниз точно так же. Обратите внимание: эти данные выдуманы для иллюстрирования принципа.

линия будет нашей статистической моделью этих данных — и с ее помощью предсказывать, сколько осадков будет выпадать каждый месяц в следующем году. Самое небрежное из возможных решений — просто провести прямую, как на рисунке 3А, только вот на имеющиеся данные это не очень-то похоже: если бы мы попробовали использовать эту прямую для прогнозирования измерений будущего года, предсказывая абсолютно одно и то же количество осадков каждый месяц, мы бы справились с заданием хуже некуда. Далее, мы могли бы использовать кривую линию, проходящую через данные так, как показано на рисунке 3Б, что было бы добротной аппроксимацией — такая кривая служила бы полезной моделью для предсказывания значений на следующий год. Однако есть риск на этом не остановиться, а провести линию через каждую отдельную точку, вихляя в разные стороны, как на рисунке 3В. Такая модель прекрасно согласуется с *нашим* набором данных, это идеальное описание имеющихся точек, но каковы шансы, что данные в следующем году будут совершенно точно так же прыгать то вверх, то вниз? Они невысоки.

Проводя линию столь близко к точкам, мы моделируем лишь случайный шум, существующий в наших данных. Это и есть чрезмерная аппроксимация.

Вот что неосознанно делают ученые, прибегающие к *p*-хакингу: придают слишком большое значение тому, что зачастую представляет собой лишь случайный шум, и считают это частью модели, а не досадным отклонением, которое ради реального сигнала (если таковой вообще существует) следует оставить без внимания. Горе тому, кто берет переподогнанную модель после *p*-хакинга и пытается применить ее к другим выборкам: она обусловлена определенными расходящимися тропками, которыми ее создатели следовали сквозь свои зашумленные данные, так что модель эта, скорее всего, мало что скажет нам о мире за пределами единичного набора данных.

Легко понять, почему ученые соблазняются переподгонкой. Если сосредоточиться только на собственных данных и забыть, что ваша работа — формулировать общие утверждения о мире, то модель вроде изображенной на рисунке 3B, чудесно описывающая данные, кажется крайне привлекательной: нет никаких неопределенностей, беспорядочных точек, уклоняющихся от проведенной вами линии. Однако столь притягательной эту модель делает не ее аккуратность как таковая, бесхитростно соединить точки на графике можно и без всякого научного знания. А как насчет статьи, звучащей так, словно вы придумали конкретную форму линии (свою теорию) *еще до* сбора данных? Теперь к вам приковано внимание научной общественности — а как мы знаем, основная цель в науке состоит в том, чтобы убедить других ученых, что вашу модель, теорию или исследование следует принимать всерьез.

Те же побуждения справедливы и для *p*-хакинга в более широком смысле: исследования, которые не омрачены случайным незначимым результатом, торчащим посреди *p*-значений, поголовно меньших 0,05, выглядят гораздо привлекательнее. Вспомним, как Стивен Джей Гулд сказал про науку: "профес-

сия, присваивающая статус и авторитет за *аккуратные и недвусмысленные открытия* [курсив мой]". Социальный психолог Роджер Джинер-Соролла согласен: "При прямой конкуренции между статьями… работе, чьи результаты все значимы и непротиворечивы, отдадут предпочтение перед той, что столь же хороша, но сообщает результат без прикрас, дабы прийти к более правомочному выводу"[75].

Здесь мы видим, как публикационное смещение и *p*-хакинг оказываются двумя проявлениями одного феномена — желания стереть результаты, которые плохо соотносятся с заранее сложившейся теорией. Этот феномен в хитроумном метанаучном исследовании обнажила группа специалистов, изучающих работы по экономике и управлению. Они воспользовались фактом, что некоторые результаты включаются в диссертации до того, как их по всей форме опишут и отправят в научные журналы для публикации. Происходящее между включением в диссертацию и опубликованием в журнале эти исследователи назвали "окукливанием". К моменту выхода в виде финальной публикации исходно неказистые наборы данных часто превращались в красивых бабочек, поскольку все неопрятно выглядящее, незначимое отбрасывалось или изменялось ради стройного изложения положительных результатов[76]. В большинстве случаев творцы сих превращений, вероятно, думали, что, устраняя некоторые результаты, позволяют своим данным четче "рассказывать историю", — и возможно, старшие коллеги учили их, что так и надо делать, дабы уверить рецензентов: статью стоит публиковать[77]. На самом же деле они обрекали будущих ученых на безнадежно предвзятую картину того, что происходило в исследовании.

Жажда красивых результатов подтачивает даже "точнейшие" из наук. Физик Сабина Хоссенфельдер в своей книге "Уродливая Вселенная" утверждает, что физики до такой степени увлеклись своими собственными построениями, сосредоточившись на элегантности и красоте моделей вроде теории струн, что на практике проверить, действительно ли эти мо-

дели верны, они не в состоянии[78]. Хотя сложнейшая математическая работа специалистов по теории струн, казалось бы, неимоверно далека от кухонной, бытовой науки Брайана Вонсинка, оба типа исследований могут пропитаться одинаковыми видами человеческих предубеждений.

Равно как и области, где из-за подобных предубеждений на кону оказываются жизни людей. Поколениям студентов-медиков рассказывали, вполне справедливо, что двойные слепые рандомизированные плацебо-контролируемые испытания — золотой стандарт для доказательства эффективности новых средств лечения. При должном проведении такие испытания вычленяют эффект плацебо, предвзятость со стороны врачей, осуществляющих лечение, ложные результаты, вызванные факторами, не имеющими отношения к испытываемому средству (так называемые спутывающие факторы), и многие другие проблемы, которые запутывают клинические исследования. Но даже строжайшим образом контролируемое клиническое испытание не может исключить предвзятость, возникающую после получения результатов, — предвзятость, расцветающую при анализе данных.

Поведение по типу техасского стрелка в медицинских исследованиях часто именуют “переключением на другие исходы”* (еще одно название для *p*-хакинга). Давайте вернемся к нашему гипотетическому исследованию разницы в росте между мужчинами и женщинами. Допустим, вы попутно зафиксировали еще какие-то факты об участниках, например их вес, количество часов, проводимых еженедельно перед телевизором, оцененный ими самими уровень стресса. Это были второстепенные показатели — любопытные, конечно, но не главные для исследования. Что произойдет, если вы не обнаружите искомую статистически значимую разницу в росте, но *обнаружите* значимое различие между мужчинами и женщинами по, скажем, времени, посвящае-

* Устоявшегося перевода этого термина на русский язык пока нет, в оригинале — *outcome-switching*.

мому просмотру телевидения? "Переключение на другие исходы" — это когда вы решаете представить работу так, словно она изначально была посвящена времени, проводимому перед телевизором. Однако у такого шага есть уже известные нам недостатки: при этом хоронится потенциально полезное знание об отсутствии разницы в росте, а тот факт, что проводились дополнительные статистические тесты, означает, что нам нужно проявить предельную осторожность при интерпретировании результатов. Если вы скрываете от читателей статьи полный объем проведенного статистического анализа, они не будут остерегаться возросшего риска ложноположительных результатов.

С 2005 года Международный комитет редакторов медицинских журналов, признавая серьезную проблему с публикационным смещением, постановил, что все клинические исследования на людях должны регистрироваться в публичном доступе прежде, чем будут проведены, — иначе в большинстве топовых медицинских журналов их не допустят до публикации[79]. Идея была в том, чтобы создать преграду для запихивания работ в "картотечный ящик", ведь все будут знать: проведены-то они были. Это постановление породило полезный побочный эффект — опубликованный список планов для каждого исследования, включая исходы, которые ученые собирались изучать[80]. Просматривая потом эти регистрационные записи, наблюдатели могут заметить нестыковки между предлагавшимся планом и написанной в итоге статьей. В проекте "Сравни испытания" *(COMPare Trials)* Бена Голдакра была предпринята попытка взять все клинические исследования, опубликованные в пяти самых престижных медицинских журналах за четыре месяца, и сопоставить с их регистрационными записями[81]. Из шестидесяти семи работ лишь в девяти сообщалось обо всем, что было заявлено. Суммарно по всем статьям триста пятьдесят четыре результата просто исчезли на пути от регистрации до публикации (есть основания полагать, что p-значения для большинства из них превышали 0,05),

тогда как триста пятьдесят семь незарегистрированных результатов появились в журнальных статьях *ex nihilo** [82]. При подобной ревизии регистрационных записей в области анестезии выяснилось, что в 92 % испытаний исследователи "переключали" как минимум один исход — и, что предсказуемо, переключение происходило скорее в сторону статистически значимых результатов[83].

Невозможно узнать наверняка, к скольким пациентам применяли бесполезные способы лечения — и давали ложную надежду — из-за клинических испытаний, подверженных *p*-хакингу, однако это число совершенно точно неимоверное[84]. Вспомним обсуждавшиеся выше метаанализы. Даже если не учитывать тот факт, что некоторые исследования часто исчезают из-за публикационного смещения, общий, совместный эффект — призванный служить надежным обобщением всего накопленного знания в данной области — окажется далек от реального, когда включенные в метаанализ работы все приукрашены *p*-хакингом[85]. Вероятно, вы задаетесь вопросом, как врачам и их пациентам верить медицинской литературе, настолько пропитанной предвзятостью, за исключением меньшинства четких, хорошо воспроизводимых результатов. Мой ответ: не имею ни малейшего представления.

Стремление публиковать привлекающие к себе внимание, неоспоримые, статистически значимые результаты — один из самых универсальных источников предубеждений в науке. Но вдобавок и другие искажающие силы оказывают свое влияние. Первое, что приходит в голову, — деньги. В США (где легко раздобыть нужные числа) чуть более трети зарегистрированных медицинских испытаний за последние годы финансировались фармацевтической промышленностью[86]. Насколько такое финансирование — компаниями, планиру-

* Из ничего *(лат.)*.

ющими продавать лекарство, если оно окажется действенным, — влияет на результаты? Метанаучные исследования, посвященные клиническим испытаниям, в целом показывают, что финансируемые индустрией проверки лекарств действительно чаще дают положительные результаты. В недавнем обзоре сообщалось, что на каждое успешное испытание, финансируемое правительством или некоммерческой организацией, приходится 1,27 успешного испытания, финансируемого фармацевтическими компаниями[87]. Возможно, эта предвзятость коренится в самом дизайне исследований: есть некоторые свидетельства того, что в испытаниях, оплачиваемых компаниями, их новое лекарство обычно сравнивается скорее с бесполезным плацебо, чем со следующим по эффективности альтернативным средством, за счет чего их новый продукт искусственно делается привлекательно выглядящим[88]. Но главным образом причина, вероятно, в тех факторах, которые мы обсуждали в этой главе: известно, например, что испытания, спонсируемые индустрией, чаще отправляются в "картотечный ящик", чем финансируемые из других источников[89].

В большинстве журналов сейчас требуется, чтобы в конце публикуемой статьи, в разделе "Конфликт интересов", авторы сообщали о любых деньгах, полученных, скажем, за консультирование фармацевтической компании[90]. Однако с иными видами финансового конфликта интересов так не обходятся. К примеру, многие ученые строят успешную карьеру на базе своих научных результатов, выпуская книги-бестселлеры и регулярно получая пяти- или шестизначные суммы за лекции, бизнес-консалтинг и речи на церемониях вручения дипломов в университетах[91]. Спору нет, люди вправе платить сколько хочется за лекторов, консультантов и контракты на написание книг. Но когда успешная карьера покоится на правдивости конкретной теории, ученый обретает новую мотивацию на своей основной работе: публиковать только те статьи, где эта теория подтверждается (либо

подвергать их *p*-хакингу до тех пор, пока это не случится). Это такой же финансовый конфликт интересов, как все другие, причем он обостряется дополнительными репутационными соображениями. Казалось бы, ученые в подобной ситуации должны прозрачности ради включать в конец каждой будущей статьи, относящейся к делу, заявление о конфликте интересов[92].

Помимо финансовых и даже репутационных интересов есть и другое предубеждение, которое редко обсуждается открыто. Это предвзятость ученого, искренне *желающего*, чтобы его исследование дало важные результаты, поскольку это сулило бы прогресс в борьбе с какой-то болезнью человека, или общества, или окружающей среды либо же иной серьезной проблемой. Дело даже не в желании получить значимые результаты ради публикации (хотя и это существенное давление, как мы увидим в последующих главах). Просто ученый имеет благие намерения и хочет чувствовать, что его работа приносит пользу. Можно было бы назвать это “предвзятостью благих намерений”. Когда исследование, которое вы спланировали для проверки своего нового средства лечения, дает отрицательные результаты, это оказывает сокрушительное действие, ведь получается, что человечество нисколько не приблизилось к тому, чтобы помочь больным людям. Приходишь в уныние, когда выдвигаешь гипотезу о связи некоего биологического фактора с определенным заболеванием — и выясняется, что шел ты по ложному следу. По крайней мере, такими могут быть чувства, если у вас неправильное отношение к науке. Ценность положительных, статистически значимых результатов в науке столь высока, что многие исследователи забывают: *отрицательные результаты тоже важны*. Тот факт, что метод лечения не работает или болезнь не связана с неким биомаркером, — это ценная информация: значит, в дальнейшем мы сможем тратить время и деньги на что-то другое. Если исследование правильно организовано, то представляет интерес вне зави-

симости от того, дает ли оно положительные *или* отрицательные результаты.

Предубеждения, которые мы обсуждали до сих пор, поражают отдельных ученых. Но помните: наука — штука социальная. Хотя распространение результатов в исследовательском сообществе способно (хотя бы частично) компенсировать предубеждения отдельных ученых, эти предубеждения могут развиться в опасное групповое мышление, когда их разделяет все сообщество. В 2019 году научная журналистка Шэрон Бегли написала шокирующую статью о противоречиях, окружающих "гипотезу амилоидного каскада" при болезни Альцгеймера. Идея в том, что накопление белка под названием "бета-амилоид", который можно наблюдать в мозге в виде "бляшек" (заметил их впервые более ста лет назад Алоизиус Альцгеймер, в чью честь и назвали этот тип деменции), является основной неврологической причиной катастрофической потери памяти и других нарушений, сопровождающих развитие болезни[93]. Бегли отмечает: несмотря на то, что в разработке средств лечения от других связанных со старением заболеваний, таких как рак и сердечно-сосудистые патологии, прорывы были, болезнь Альцгеймера упорно остается неизлечимой, поскольку клинические исследования лекарств, призванных облегчить симптомы благодаря разрушению амилоида, одно за другим приносили лишь разочарование[94]. Почему так? По словам нескольких ученых, с которыми Бегли общалась, дело попросту в том, что амилоидная гипотеза неверна. Хотя амилоидные бляшки *связаны* с симптомами, не они их вызывают: борьба с амилоидом не поможет вылечить болезнь[95].

Несогласные исследователи описывали, как приверженцы амилоидной гипотезы, среди которых много влиятельных, солидных профессоров, действовали как "заговорщики", уничтожая статьи, ставящие гипотезу под сомнение, разгромными рецензиями и подрывая попытки инакомыслящих ученых найти финансирование и должности. Бегли полагает,

что происходит это не обязательно по какому-то сознательному решению: просто апологеты амилоидной гипотезы так искренне в нее верят, считая оптимальным путем к прорыву в лечении болезни Альцгеймера, что у них создалось сильнейшее предубеждение в ее пользу.

Можно даже осмелиться предположить, как сделали некоторые из опрошенных Бегли исследователей, что мы бы уже продвинулись в лечении (или искоренении) болезни Альцгеймера, отбрось ученые амилоидную гипотезу много лет назад. Возможно, проблема просто в том, что это заболевание, поражающее как-никак мозг, самый сложный орган из всех, особенно неприятное — или что в опубликованных на сегодняшний день клинических испытаниях лекарств, разрушающих амилоид, неверно ставился вопрос (скажем, в них участвовали люди уже слишком пожилые, кому амилоид успел причинить вред)[96]. Однако случаи травли и принуждения, происходящие, когда исследователи ставят амилоидную гипотезу под сомнение, намекают, что это область, где предвзятость стала общей, новые идеи не получают того внимания, какого заслуживают, и где ученые раз за разом не удосуживаются применить принцип организованного скептицизма к своим собственным излюбленным теориям.

А что насчет исследователей, движимых скорее идеологической или политической заинтересованностью в правдивости или ложности результатов? Один из самых примечательных разделов о конфликте интересов встретился мне в статье, посвященной здравоохранению, про так называемый эффект Глазго. Так окрестили явление, заключающееся в том, что жители Глазго — и в целом Шотландии — умирают в среднем раньше, чем жители других похожих городов или стран, даже при учете уровня бедности и лишений. В статье после описания подтверждений этого явления утверждалось, что корень уникальной проблемы — "политическая атака" на Шотландию консервативного правительства премьер-министра Маргарет Тэтчер 1980-х годов, с мерами по деиндустриализации

и ослаблению влияния профсоюзов. В соответствующем разделе публикации не числятся финансовые конфликты интересов, зато указано, что ведущий автор, Герри Маккартни, — "член Шотландской социалистической партии"[97]. Похвальная, скажу я вам, и редкая честность[98].

Моя собственная область, психология, не чужда ученым, причисляющим себя к левым. Более того, перекос в психологии весьма силен: опросы в США выявили, что соотношение либералов и консерваторов приблизительно равно десяти к одному. (В других областях, например в инженерном деле, бизнес-исследованиях, информатике, такого не наблюдают — там, как и в Америке в целом, соотношение примерно равное.) Психология — изучающая людей и их поведение — обычно гораздо крепче сращена с политическими соображениями, чем, скажем, теоретическая физика или органическая химия. В статье 2015 года Хосе Дуарте и несколько других видных психологов писали, что частично по этой причине политические предубеждения могут особенно вредить их области науки[99]. Они утверждали, что, как и в примере выше с групповым мышлением, если подавляющее большинство в сообществе разделяет одинаковые политические взгляды, важная функция рецензирования — подвергать любые заявления суровейшей проверке из возможных — существенно ослабляется. Мало того, могут сместиться приоритеты в вопросах, касающихся тем исследования вообще: ученые станут уделять непропорционально много внимания каким-то политически приемлемым темам, даже если они подкрепляются довольно слабыми доказательствами, и избегать тех, что идут вразрез с определенной установкой, даже если подкрепляются они доказательствами надежными[100].

Критики уклона психологии в либерализм обрушились, например, на так называемую *угрозу подтверждения стереотипа*[101]. Согласно этой идее, девочки хуже пишут тесты по математике, когда им напоминают о стереотипе, будто мальчики в математике сильнее. Такая идея показалась бы

гораздо более правдоподобной человеку с социально-либеральными политическими взглядами, при которых стереотипы и сексистские предрассудки считаются мощными силами, воздействующими на личность людей и формирующими общество. Доказательства этого феномена довольно слабые и, наверное, подвержены публикационному смещению, поскольку в метаанализе 2015 года, когда изучались все релевантные исследования, посвященные угрозе подтверждения стереотипа, обнаружилась явная проплешина, где должны были бы оказаться небольшие исследования на эту тему, давшие отрицательные результаты — то есть показавшие, что девочки одинаково хороши в математике при упоминании стереотипов и без оного (подразумевается, что воронкообразная диаграмма выглядела подобно нашей с рисунка 2Б)[102]. Весьма вероятно, что небольшие исследования с отрицательными результатами были проведены, а затем отброшены, ведь они не соответствовали преимущественно либеральным предубеждениям ученых, искажая нашу доказательную базу по этому важному образовательному вопросу[103]. Однако как деформированная воронкообразная диаграмма еще не обязательно обличает предвзятость, так и публикационное смещение само по себе не следует считать прямым доказательством политических предубеждений. В конце концов, мы знаем, что ученые в принципе предпочитают положительные результаты отрицательным, независимо от политических взглядов. Тем не менее дискуссия об угрозе подтверждения стереотипа лишний раз предостерегает нас от возможной опасности: общие предположения, политические или иные, иногда могут сдерживать самокритику, необходимую для прогресса науки.

Еще дискуссия об угрозе подтверждения стереотипа подводит нас к проблеме сексизма — одного из самых обсуждаемых предубеждений среди ученых (наряду с политическими). Важная тема, касающаяся сексизма в науке, — насколько женщины представлены в разных научных областях и какие по-

зиции они там занимают в иерархии. Но еще обсуждается влияние сексистских предубеждений на то, как вообще наука делается[104]. Например, специалисты по нейронаукам, работающие с животными вроде мышей, обычно экспериментируют только над самцами изучаемого вида. А все потому, что самки часто считаются более подверженными гормональным изменениям, которые служат трудно контролируемым и способным исказить результаты источником вариабельности в вопросах, касающихся мозга животных и их эмоций. Причем тот факт, что у самцов уровень гормонов, к примеру тестостерона, тоже очень сильно скачет (обычно в соответствии с их статусом в иерархии подчинения), игнорируется. В итоге потенциально порождается неменьшая вариабельность в поведении[105]. Если в силу этого предубеждения ученые исследуют лишь самцов, результаты могут оказаться не универсально применимыми, ведь самки действительно отличаются от самцов во многом, что касается мозга и поведения[106]. Как отметила специалист по нейронаукам Ребекка Шански, "заболеваниями вроде клинической депрессии и посттравматического стрессового расстройства вдвое чаще страдают женщины, однако типичные поведенческие тесты, призванные смоделировать соответствующие симптомы у грызунов, разработаны и валидированы на самцах"[107].

Психолог и философ Корделия Файн, проницательный критик низкокачественных, зачастую отмеченных *p*-хакингом исследований, в которых ученые пытаются объяснить поведенческие различия между полами, просто-напросто связывая их с уровнем тестостерона и между делом забывая о социальной подоплеке, тоже занималась проблемой возведения самцов в ранг участников медицинских исследований "по умолчанию" и низведения самок до некой вторичной категории или даже аберрации, отклонения[108]. В своей заметке 2018 года для журнала *The Lancet* Файн предположила, что "феминистская наука" может восстановить баланс, подсвечивая подобные упущения. Она допускает, что кто-то будет

настроен скептически: “Обычно считается, что, хотя феминизм неплох для гендерных исследований, его не следует подпускать к науке, иначе политические предпочтения, связанные с тем, какими должны быть женщины, мужчины и наш мир, исказят научные свидетельства того, каковы они на самом деле”[109]. Но разве, спрашивает она, мы все так или иначе не пристрастны? Ведь “каждый имеет точку зрения, никто не отличается «взглядом из ниоткуда». Открывая путь феминистской науке, мы не распахиваем дверь перед политическими предубеждениями. Это значит только, что мы не все смотрим сквозь одно и то же мутное стекло”[110].

В каком-то смысле это политическое отражение проблемы, сформулированной Хосе Дуарте и его коллегами, о малом числе консерваторов в области психологических исследований. Хотя Дуарте и Файн расставляют разные акценты, оба они ратуют за бóльшую представленность в науке недостаточно представленных взглядов. И это очень важно: при обсуждении научных данных один из самых значимых аспектов — то, что люди смотрят на вопрос под разными углами. Но призывать ученых — или науку в целом — принимать какой-либо набор социально-политических обязательств, по моему мнению, неразумно, даже если в краткосрочной перспективе это поможет решить какие-то конкретные проблемы. Вместо этого нам нужно постараться минимизировать негативное влияние собственных предубеждений на принимаемые нами научные решения и на проводимый нами научный анализ. (Хотя кто-то, например Файн, мог бы возразить, что это неосуществимо практически, а то и невозможно, в следующих главах мы обсудим некоторые идеи, как это можно сделать.)

Попытка скорректировать предвзятость в науке внесением равной, но противоположной по знаку дозы предвзятости лишь осложняет проблему и, вероятно, запускает порочный цикл усугубления разлада между идеологическими лагерями. Мало того, предложение, чтобы ученые гордо позволяли своим идеологическим взглядам влиять на исследования, на-

рушает мертоновскую норму как бескорыстности (ведь при этом ненаучным соображениям дозволяется посягать на исследования), так и универсализма (поскольку к научным аргументам могут предъявлять разные требования в зависимости от политических пристрастий высказавшего их ученого). Причина, по которой Герри Маккартни в статье о Маргарет Тэтчер решил заявить о своем членстве в социалистической партии, заключается в том, что иначе, всплыви этот факт после публикации, он скомпрометировал бы автора — учитывая научные выводы статьи, хорошо согласующиеся с идеологией соцпартии. Даже когда политические взгляды не столь прямо связаны с результатами исследования, как в этом случае, ученым следует не менее щепетильно относиться к возможным пересечениям их политических убеждений с научной работой.

Идея, согласно которой всякий ученый имеет идеологические представления, влияющие на его работу, замыкает круг, возвращая нас к истории измерения черепов Сэмюэлем Мортоном, а также к критике его предубеждений Стивеном Джеем Гулдом. В 2011 году антрополог Джейсон Льюис и его коллеги вернулись не просто к вычислениям Мортона, как сделал Гулд, но к самим черепам из мортоновской коллекции, хранящейся в Пенсильванском университете, перемерив примерно половину из них по современным методикам[111]. Льюис и его коллеги согласились, что ранжирование Мортоном разных групп людей было откровенно расистским, и подтвердили, что он действительно совершал измерительные ошибки. Однако, настаивали они, ошибки не были систематическими в том смысле, о котором говорил Гулд, — неправильные измерения относились скорее ко множеству черепов и не встречались в одной расовой группе чаще, чем в другой. Еще огрехи могли происходить по вине ассистента, о чьих ошибках Мортон упоминал, а не из-за “вероятного

сценария", по которому Мортон, согласно Гулду, заталкивал больше семян в черепа белых.

Более того, Льюис и его коллеги утверждали, что Мортон вовсе не ловчил с группами (забывая упомянуть о группах небелых людей с большим средним размером черепа), в чем обвинял его Гулд. Они заявили, что на самом деле Гулд допустил *собственные* ошибки, разбивая мортоновскую выборку на подгруппы таким образом, который согласовывался с его излюбленными представлениями о равенстве размеров черепов. В предисловии к своей книге "Неверное измерение человека" *(The Mismeasure of Man)* Гулд открыто признался в сильной приверженности идеям социальной справедливости и либеральной политики[112]. В статье Льюиса делался такой вывод: "по иронии анализ работы Мортона самим Гулдом является едва ли не более наглядным примером того, как предубеждения влияют на результаты"[113].

Провокационное высказывание. Могло ли это быть, чтобы легендарная работа такого высокоавторитетного человека, как Гулд, оказалась настолько ошибочной? Не все считали, что история с Льюисом и его коллегами проста. Философ Майкл Вайсберг полагал, что главный посыл гулдовских рассуждений остался верным, хотя и признал, что новые измерения черепов корректны, и согласился, что Гулд в своем анализе кое-где напортачил[114]. Ведь то, что ассистент мог случайно допустить какие-то ошибки, было лишь подозрением; имеющиеся доказательства все еще не противоречили сценарию, согласно которому Мортон (или, возможно, его ассистент) был предубежден против приписывания небелым бо́льших размеров черепа. И после исправления ошибок разница в размерах черепов у представителей различных рас все равно была очень мала, а в этом и заключалась суть критики Гулда. Финальный (на сегодняшний день) аккорд прозвучал в 2018 году, когда обнаружились некоторые дополнительные измерения черепов, сделанные самим Мортоном. С учетом этих новых данных предположение, будто

расхождения в измерениях с помощью семян и дроби у Мортона были значительнее для расовых групп, не пользующихся его благосклонностью, что составляло весомую часть гулдовских доводов о его предвзятости, стало выглядеть безосновательным[115].

Конечно, если иметь в виду поиск ответа на какой-либо содержательный научный вопрос, то все эти баталии над запыленными черепами бессмысленны. Даже если бы мы допустили, что сомнительная идея Мортона о связи между размерами черепов и "умственными и нравственными способностями" их обладателей из разных групп верна, его коллекция не является репрезентативной выборкой черепов со всего мира, а значит, по ней очень мало что можно заключить — а то и вообще ничего — о любых общих различиях между группами[116]. Эти прения имеют, однако, четкий смысл в контексте предубеждений в науке: за наблюдателями тоже должен кто-то наблюдать, а разоблачителей — разоблачать. И даже тогда стоит проверять, а разобрались ли те, кто разоблачал разоблачителей, что к чему. Как мы теперь понимаем, каждый человек подвержен каким-то своим предубеждениям; только в мортоновской эпопее мы видели примеры того, что можно было бы назвать "расовой предвзятостью", "эгалитаристской предвзятостью" и "предвзятостью, связанной с желанием доказать, что известный ученый прошлого был неправ", и все они наверняка способствовали искажению истины[117].

Предубеждения — неизбывная часть человеческой природы, и наивно было бы думать, что их возможно вытравить из какой-либо области деятельности. Тем не менее у нас имеются инструменты, позволяющие добиться чуть большей объективности. Статистический анализ как таковой призван снять с предвзятого человека бремя принятия решений — и все же мы видели, как легко подправить числа в свою пользу. Рецен-

зирование призвано служить преградой для наших предрассудков — однако мы наблюдали, как попытки убедить рецензентов и редакторов опубликовать нашу работу приводят к тому, что неудобные результаты либо вообще заметаются под ковер, либо же насильственно приводятся в согласие с нашим предвзятым мнением. Любое из предубеждений — вызванное ли научным догматизмом, политической позицией, финансовым давлением или просто жаждой получить статистически значимые результаты — может быть совершенно неосознаваемым. Как раз неосознаваемость и наделяет их силой: если вы хотите убедить рецензентов и весь мир в своей правоте, полезно сначала убедить самого себя. А еще из-за своей неосознаваемости предубеждения так пугающи.

Беспристрастность, как и честность, — одно из качеств, которые должны быть присущи науке. Горькая правда в том, что зачастую, занимаясь наукой, мы порождаем полную их противоположность. В следующей главе мы добавим еще один пункт к нашему списку парадоксов: хотя наука нацелена на верное понимание фактов, наши исследования частенько изобилуют самыми банальными видами ошибок.

Глава 5
Недобросовестность

> Неведение — это чистый лист, на котором мы можем писать, но заблуждение — лист замаранный, с которого мы прежде должны все стереть.
>
> Чарльз Калеб Колтон *“Лакон, или Многое в немногих словах” (1820)*[1]

В физике есть законы, в математике — доказательства, а в социальных науках есть “стилизованные факты”, утверждения вроде таких: “люди более образованные, как правило, получают больший доход” или “демократии не вступают в войны друг с другом”[2]. Они не столь незыблемы, как законы и доказательства, но призваны выражать простым языком общие, важные и воспроизводимые результаты. Как физики желали бы открыть новый закон (или способ нарушить те, что уже известны), а математики без конца трудятся, чтобы доказывать свои теоремы, многие специалисты по социальным наукам, особенно экономисты, стремятся обнаружить стилизованный факт, который мог бы ассоциироваться с их именем — и который люди, принимающие важные решения, могли бы легко запомнить. Когда в 2010 году экономисты Кармен Рейнхарт и Кеннет Рогофф опубликовали свою масштабную статью, они думали, что сорвали джекпот.

Два года политики лихорадочно пытались разобраться с последствиями финансового кризиса, грянувшего в 2008-м, и с наступившей после него Великой рецессией. На фоне всевозможных противоречивых советов статья Рейнхарт и Рогоффа, озаглавленная “Рост во времена долга”, стала спасением: в ней убедительно обосновывалось, что рекомендован единственный и конкретный метод экономических действий — жесткая экономия[3].

Рейнхарт и Рогофф изучали отношение долга к ВВП — соотношение между тем, что страна должна своим кредиторам (государственным долгом, известным еще как "суверенный долг"), и тем, какие новые товары и услуги она способна произвести (валовым внутренним продуктом). Они исследовали связь между этим отношением и темпами роста экономики в стране. На основании исторических данных по нескольким десяткам стран они показали, что, когда отношение долга к ВВП было низким — скажем, если долг составлял от 30 до 60 % от ВВП, — особой связи с ростом экономики не наблюдалось, зато в странах, где отношение превышало вполне определенный порог в 90 %, экономика сокращалась.

Вероятно, потому, что основной результат так легко поддавался превращению в стилизованный факт ("отношение долга к ВВП выше 90 % — беда для экономического роста"), исследование оказало чрезвычайно большое влияние. Оно не только широко освещалось в средствах массовой информации, но и помогло сформировать во многих государствах политику жесткой экономии в ответ на рецессию: идея была в том, что правительства должны попытаться выплатить свои долги (сокращая расходы, повышая налоги или и то и другое), чтобы опустить этот показатель ниже критической отметки в 90 %. Исследование было прямо упомянуто в важной речи Джорджа Осборна, занимавшего тогда пост министра финансов Великобритании, и в заявлении республиканских членов сената США и бюджетной комиссии палаты представителей[4]. По мнению экономиста Пола Кругмана, критикующего идею Рейнхарт и Рогоффа, так много политиков, выступающих за жесткую экономию, ссылались на их исследование, что оно "оказало, вероятно, большее непосредственное влияние на общественную дискуссию, чем любая другая статья в истории экономики до этого"[5].

Тем большее беспокойство вызывают дальнейшие события. В 2013 году критики исследования обнаружили ошибку в электронной таблице *Microsoft Excel*, использовавшейся

Рейнхарт и Рогоффом при анализе данных: долг нескольких стран не был учтен[6]. А именно: в таблице при расчетах не был учтен долг Австралии, Австрии, Бельгии, Канады и Дании. Ужасающе банальная причина? Опечатка. Когда ее устранили, а также внесли поправки к еще парочке спорных аналитических решений, принятых Рейнхарт и Рогоффом, соотношение между коэффициентом задолженности и экономическим ростом резко изменилось[7]. В статье говорилось, что средний темп роста при отношении долга к ВВП выше 90 % составлял −0,1 %; после внесения поправок он составил +2,2 %. Все же ничего магического в этом пороговом значении 90 % не было — выше него темпы роста не стали внезапно отрицательными. На самом же деле существовал "широкий диапазон показателей роста ВВП при любом уровне государственного долга"[8]. Если бы в статье исходно содержалось подобное осторожное утверждение, гораздо более сложное, чем соответствующий стилизованный факт, вряд ли она привлекла бы столько внимания.

Итак, изменила ли опечатка мировую экономику? Не очень-то. Хотя статья и порожденный ею стилизованный "факт" имели необычайно широкое влияние, доводы за удерживание отношения долга к ВВП на низком уровне не упираются в единственное исследование[9]. Обнаружение той опечатки лишь ослабило выводы Рейнхарт и Рогоффа, а не сделало совершенно несостоятельными. И, как отмечалось выше, критики сосредоточились не *только* на опечатке. Мы не можем переиграть мировую политику, чтобы выяснить, нашлись бы у тех, кто ратует за меры жесткой экономии, другие причины для их введения, если бы исследования Рейнхарт и Рогоффа никогда не существовало, — без сомнения, нашлось бы немало. И все же то обстоятельство, что столь примитивная ошибка в итоге попала на стол к влиятельным политикам, должно вызывать глубокое беспокойство, не в последнюю очередь потому, что неизбежно заставляет нас задуматься: а сколько *еще* там должно быть подобных ошибок,

компрометирующих научную литературу и, кто знает, даже влияющих на реальные решения.

Ответ: их слишком много. В этой главе рассматривается два вида недобросовестности в науке. С первым мы только что столкнулись: непроизвольные ошибки, вносимые в научный анализ по невнимательности, недосмотру или небрежности. Второй — это когда ученые, которым положено понимать, что к чему, закладывают ошибки в сам способ проведения исследований, что может объясняться неподготовленностью, безразличным отношением, забывчивостью или, как бы жестоко это ни звучало, полнейшей некомпетентностью. Подобные ошибки из-за недобросовестности служат еще одним болезненным свидетельством того, что наша научная система не справляется со своей главной задачей, ради выполнения которой и была задумана.

Насколько в научных работах распространены числовые ошибки? В 2016 году группа голландских исследователей под руководством психолога Мишель Нейтен попыталась это выяснить. Они представили алгоритм под названием *statcheck*, своего рода “спелл-чекер для статистики”[10]. Когда вы вводите в *statcheck* научную статью, он проходится в ней по всем числам и отмечает ошибки в *p*-значениях. Алгоритм способен это сделать, поскольку многие числа в статистических тестах зависят друг от друга, поэтому, если известны лишь некоторые из них, всегда можно воспроизвести остальные (как благодаря теореме Пифагора вы всегда можете вычислить гипотенузу треугольника, если известны длины двух других сторон). Если *p*-значение и другие связанные с ним числа друг с другом не согласуются, что-то, вероятно, не так. Нейтен и ее коллеги прогнали через *statcheck* более тридцати тысяч статей — гигантскую выборку исследований, опубликованных в восьми главных журналах по психологии с 1985 по 2013 год[11]. От того, что они обнаружили, становится просто неловко.

Почти в половине работ, содержащих нужные статистические данные, имелось как минимум одно числовое несоответствие. Справедливости ради отметим, что большинство ошибок были незначительными и практически не затрагивали общие результаты. Однако некоторые несоответствия сильно повлияли на выводы исследования: в 13 % случаев были допущены серьезные ошибки в стиле Рейнхарт и Рогоффа, которые могли полностью изменить интерпретацию результатов (например, превратить статистически значимое *p*-значение в незначимое или наоборот). Конечно, эти несоответствия могли возникнуть по разным причинам: от простых опечаток и ошибок копирования вплоть до сознательного мошенничества. *Statcheck* — это лишь способ подсветить ошибки в научном тексте, а не установить причины их появления.

Один из самых интересных результатов анализа Нёйтен показывает, как недобросовестность сопряжена с предвзятостью. Несоответствия, отмеченные алгоритмом *statcheck*, были, как правило, в пользу авторов — то есть ошибочные числа делали результаты более, а не менее согласующимися с гипотезой исследования. Будь это просто абсолютно случайные опечатки, нельзя было бы ожидать, что в среднем они сдвинут результаты в какую-то одну сторону. Похоже, однако, как мы могли бы предсказать на основании своих знаний о предвзятости, будто ученые чаще пересматривали результаты, когда те отклонялись от намеченного пути. А вот ошибочные результаты, *подтверждавшие* их теорию, были попросту слишком хороши, чтобы их проверять.

Другой особенно изящный метод узнать, верны ли приведенные в статье числа, имеет явно неизящное название “тест на несогласованность средних, связанный с гранулярностью”, сокращенно он именуется тестом GRIM *(Granularity-Related Inconsistency of Means)*[12]. С помощью этого теста, разработанного расследователями в мире данных Ником Брауном и Джеймсом Хизерсом, проверяется, имеет ли смысл среднее значение (а именно — среднее арифметическое) набора чисел

с учетом того, сколько чисел в нем содержится. Представьте, что вы просите людей оценить, насколько они довольны своей работой, по шкале от 0 до 10 (и ответ дозволяется давать только в целых числах: скажем, "4" или "5", но не "3,7"). Допустим, в простейшем случае вы задали этот вопрос только двум людям и сообщаете среднее значение их оценок, то есть складываете два их результата и делите сумму пополам. Если в получившемся числе посмотреть на цифры после запятой, то возможных вариантов для них будет всего ничего: для двух человек среднее значение их ответов может заканчиваться только на ",00" или ",50". Если же у вас получилось, например, 4,40, значит, что-то точно пошло не так: при делении целого числа на два получить такую дробь невозможно.

В тесте GRIM та же логика применяется к выборкам покрупнее. К примеру, если двадцать участников оценили что-либо по шкале целых чисел от 0 до 10, вам никак не получить среднее, равное 3,08. При делении на двадцать значения после запятой могут идти только с шагом 0,05: среднее, равное 3,00, или 3,10, или 3,15, получить возможно, а 3,08 — невозможно[13]. Браун и Хизерс использовали тест GRIM, чтобы проверить выборку из семидесяти одной опубликованной статьи по психологии, и обнаружили, что в половине из них приводилось по меньшей мере одно невероятное число, а в 20 % содержалось даже несколько. Как и в случае с алгоритмом *statcheck*, ошибки, выявленные методом GRIM, могут объясняться причинами безобидными, однако же служат красными флагами, сигнализирующими о необходимости дальнейшего анализа.

Значение 3,08 в моем примере неслучайно, я выбрал его, поскольку оно сыграло свою роль в истории теста GRIM и психологических исследований в целом. В 2016 году психолог Матти Хейно применил метод GRIM для анализа одной из самых знаменитых статей по психологии всех времен — работы Леона Фестингера и Джеймса Карлсмита о "когнитивном диссонансе", вышедшей в 1959-м[14]. Сейчас эта идея ши-

роко известна: когда человека заставляют говорить или делать то, что противоречит его истинным убеждениям, он чувствует психологический дискомфорт и старается подладить эти свои убеждения под слова и действия, к которым его принудили. Участников исследования 1959 года заставляли выполнять нудные и бессмысленные задания, например без конца проворачивать штырьки в перфорированной панели. После некоторым выплачивали доллар, чтобы они говорили следующим по очереди участникам, будто задания показались им весьма интересными и забавными. Позднее при опросе участники, которым за ложь о задании заплатили, отзывались о нем как о гораздо более интересном, чем те, кому доллар не достался. Иными словами, они уменьшали свой диссонанс, заставляя себя поверить, что им было весело[15]. Увы, примененный Хейно тест GRIM продемонстрировал, что не только впечатления участников были рассогласованы — но и числа у Фестингера и Карлсмита[16]. Для выборки из двадцати человек, дающих оценку по шкале целых чисел от 0 до 10, они заявили о среднем, равном 3,08, что невозможно, как мы обсудили выше, не говоря уже о том, что тест не прошли и еще несколько средних.

Когнитивный диссонанс — исключительно полезная идея, интуитивно нам понятная, и эксперимент был хитрым и запоминающимся. Но стали бы тысячи ученых, на протяжении многих лет ссылавшиеся на статью Фестингера и Карлсмита, делать это, знай они, что она кишит неправдоподобными числами?[17] Эта история еще раз напоминает нам, что даже "классические" результаты из научной литературы — те, что, хотелось бы надеяться, проверялись самым строгим образом, — могут быть совершенно ненадежными, когда числа и другие данные, которым следует играть важнейшую роль, выступают в качестве чистых декораций — фона для захватывающей истории.

Еще числовые ошибки настораживающе распространены в научных областях, где ставки куда выше. Вспомните: самый

плодовитый научный мошенник в мире (во всяком случае на момент написания этих строк) — анестезиолог Ёситака Фудзии*. Его затяжному марафону по подделыванию данных положил конец анестезиолог Джон Карлайл, разработавший статистический метод для проверки, действительно ли рандомизированные клинические испытания рандомизированы[18]. Рандомизация по своей сути — это словно подбрасывание монеты для каждого из участников, чтобы распределить их по группам (скажем, активного препарата либо же плацебо) случайным образом, а не каким-либо заранее спланированным способом, который может быть источником предвзятости. Это процесс первостепенной важности: он призван на момент начала испытаний гарантировать, что между группами нет существенных различий. Если люди в одной группе здоровее, образованнее, старше или заметно отличаются по любому другому показателю, который может повлиять на результаты, исследование не будет честным[19]. Поэтому, если на начало рандомизированного контролируемого исследования между группами *имеются* большие различия, налицо проблема: процесс рандомизации провален. С другой же стороны, если группы подобраны *идеально* и необъяснимым образом удалось избежать железного правила о зашумленности чисел, это тоже сомнительно: даже после рандомизации между группами все равно должны быть мельчайшие различия, просто по случайности. Вот на чем основан метод Карлайла. Когда он проверил статьи Фудзии, то обнаружил данные совершенно неправдоподобно стройные: так, например, распределения заявленных возраста, роста и веса пациентов Фудзии были почти идеально синхронизированы. Шансы, что такое произойдет в реальности, меньше одного к десяти в тридцать третьей степени (то есть одного к миллиарду триллионов триллионов)[20]. Само собой, оказалось, что Фудзии — мошенник.

* Как уже отмечалось, с тех пор он был отодвинут на второе место. Однако номер один на текущий момент, Йоахим Больдт, тоже анестезиолог.

В 2017 году Карлайл применил свой обнаруживающий ошибки метод к пяти тысячам восьмидесяти семи медицинским исследованиям из восьми журналов, опять проверяя рандомизацию, которая была либо негодной, либо подозрительно совершенной[21]. Не исключено, конечно, что *некоторые* испытания выглядят сомнительно просто по невезению. Но даже принимая это во внимание, Карлайл обнаружил, что 5 % исследований содержат подозрительные данные: таким образом, он выявил сотни работ, которые, возможно, полностью искажены — а их результаты бессмысленны — из-за неправильной рандомизации групп. Лишь за малой долей этих проваленных испытаний стояло мошенничество в стиле Фудзии; судя по всему, Карлайл обнаружил в основном "невинные" ошибки. Хотя, учитывая, что стоит на кону в медицинских исследованиях — ведь врачи используют их результаты при подборе лечения для своих пациентов, — эти "невинные" ошибки могут обернуться большой бедой[22].

Самое замечательное в алгоритме *statcheck*, тесте GRIM и методе Карлайла — что их можно применить, используя лишь итоговые данные, которые обычно предоставляются в статьях, вроде *p*-значений, средних, размеров выборок и стандартных отклонений. Не требуется доступа к таблицам полных исходных данных. Пожалуй, оно и к лучшему, поскольку ученые славятся нежеланием делиться своими данными, даже когда другие *добросовестные* исследователи любезно их об этом просят. Исследование, проведенное в 2006 году, показало, что жалкие 26 % психологов согласны были отправить свои данные другим ученым по запросу через электронную почту; похожие удручающие цифры получены и в других областях. Еще ваша вероятность получить доступ к данным будет становиться тем меньше, чем больше времени проходит с момента исследования[23]. Подобное нежелание делиться данными блокирует жизненно важные процессы самокритики — все те же мертоновские нормы коллективизма и организованного скептицизма, — что лежат в основе науки.

И какими бы хитрыми ни были три перечисленных выше метода, они меркнут в сравнении со всеобъемлющей проверкой, которую можно было бы провести, имея в своем распоряжении полный, подробный набор данных. Впрочем, сейчас стремление сохранить данные в тайне (а еще, возможно, страх, что кто-то найдет в вашей опубликованной работе ошибки) явно перевешивает мертоновские мотивы для обмена ими.

В сущности, все научные направления страдают от числовых огрехов. Причем в некоторых областях имеются свои собственные типы ошибок. Рассмотрим, например, клеточные линии — фактически бессмертные клеточные культуры, которые размножаются неограниченно долгое время и используются как модели для изучения различных видов клеток, как здоровых, так и раковых. В 1958 году, через несколько лет после создания первой бессмертной клеточной линии, было замечено, что клетки из разных линий — более того, от разных биологических видов — иногда смешиваются, если ученые работали недостаточно внимательно[24]. Произойти это могло легко: кто-то неправильно подписал образец, или несколько исследователей работали бок о бок либо же в захламленной лаборатории, или, скажем, оборудование не очищается и не стерилизуется должным образом. Контаминация — новость особенно плохая, ведь некоторые клеточные линии растут быстрее и упорнее, чем другие, поэтому не успеешь и глазом моргнуть, как одна полностью заменит другую. Стоит ли говорить, что проведение эксперимента на неправильном типе клеток может полностью подорвать его результаты: вы думали, что исследуете рак кости, а используете, оказывается, клетки рака толстой кишки; вы считали, что проводите исследование на человеческих клетках, но выясняется, что используемые вами получены от свиней или крыс[25]. Вряд ли стоит удивляться, если результаты таких экспериментов нельзя будет воспроизвести. В насто-

раживающей редакционной статье из престижного журнала *Nature Reviews Cancer* говорится прямо: “Научное сообщество не сумело решить эту проблему, и поэтому опубликованы тысячи вводящих в заблуждение и потенциально ошибочных статей об исследованиях с использованием неправильно идентифицированных клеточных линий”[26].

Сколько тысяч? В 2017 году при прочесывании научной литературы в поисках исследований, где использовались известные неверно идентифицированные клеточные линии, обнаружились ошеломительные тридцать две тысячи семьсот пятьдесят пять работ с участием клеток-самозванок и более полумиллиона статей, ссылавшихся на эти “контаминированные” работы[27]. (Поскольку среди клеточных линий, изучаемых учеными, очень много раковых, областью исследований с наибольшим количеством “контаминированных” публикаций, что неудивительно, оказалась онкология.) Отдельное исследование, посвященное китайским лабораториям, показало, что были неправильно идентифицированы до 46 % клеточных линий[28]. В другом исследовании отмечается, что до 85 % клеточных линий, считавшихся недавно созданными в Китае, контаминированы клетками из США; лаборатории много и часто обмениваются клеточными линиями, поэтому такого рода ошибки быстро усугубляются[29]. Однако это проблема далеко не только Китая: было также установлено, что 36 % опубликованных исследований, в которых использовались контаминированные клетки, — из США. И множатся не только клеточные линии — множатся и научные исследования, в которых совершается такая ошибка, причем их число, вопреки тому, что ученые давно знают об этой проблеме, все растет и растет[30].

И это в неправильной идентификации клеточных линий удручает больше всего: абсолютно расслабленное отношение научного сообщества на протяжении более полувека к такой насущной проблеме, из-за которой тормозятся медицинские исследования и выбрасываются на испорченные экс-

перименты огромные суммы денег. Через два с лишним десятилетия после того, как клетки-самозванки были впервые выявлены, журнал *Nature* опубликовал редакционную статью — в ответ на случай, когда клетки обезьяны дурукули перепутали с человеческими, — где преуменьшалась значимость проблемы, а разоблачители, обнаружившие контаминацию, пренебрежительно именовались "самопровозглашенными борцами за справедливость"[31]. Конечно, ошибочная идентификация клеток не всегда сильно вредит исследованию: можно перепутать, скажем, различные виды клеток рака предстательной железы человека и все равно получить стоящие результаты. Однако часто результаты исследования становятся совершенно бесполезными, в чем можно убедиться, взглянув на все пополняющийся список статей, отозванных из-за неверной идентификации клеток[32].

Поразительно, что проблема смешения клеточных линий продолжает преследовать клеточную биологию. Изучив научную литературу только за последнее десятилетие, я нашел редакционные статьи и призывы к действию, имеющие отношение к неправильной идентификации клеточных линий, в известных журналах за 2010, 2012, 2015, 2017 и 2018 годы, а восходят эти призывы к 1950-м[33]. Сейчас благодаря новым ДНК-технологиям существуют более надежные и дешевые способы, позволяющие проверить подлинность каждой клеточной линии, поэтому в будущем, вероятно, удастся предотвращать больше подобных ошибок[34]. Однако вялая реакция специалистов по клеточной биологии в прошлом означает, что обольщаться, похоже, не стоит. Это не просто какой-то единичный случай с досадной ошибкой, а отрицание серьезного провала в течение десятилетий. Если сравнить это с происходящим в других областях — таких, например, как авиация, где колоссальные усилия направлены на расследование причин каждой авиакатастрофы, с тем чтобы выявленная проблема никогда больше не возникла, — то ученые выглядят и впрямь крайне неадекватно. И поскольку их упрямство за-

медляет исследования рака, все дело принимает очень неприятный в моральном отношении оборот.

Такой моральный аргумент, что ошибки в науке — это гораздо больше, чем просто академический вопрос, из-за вреда, который они могут нанести, в равной степени относится и к тем областям исследований, где непосредственно жертвуют жизнями. Я имею в виду, конечно, исследования на животных, где испытуемые в рамках эксперимента часто подвергаются “эвтаназии”, то есть их убивают (например, чтобы изучить их мозг после введения какого-то нового лекарства). Подобные исследования обычно строго регулируются государственными органами, поскольку практически все согласны с тем, что было бы аморально убивать лабораторных животных или даже просто заставлять их страдать, не имея на то веских научных оснований[35]. Таким образом, ученые, проводящие исследования на животных, несут на себе не только обычное научное бремя — попытаться получить точные, воспроизводимые результаты, не расточая при этом ресурсы. На них лежит еще и дополнительная ответственность: следить за тем, чтобы ошибки в постановке эксперимента и при анализе результатов не сделали бессмысленными смерть и всю ту боль, которые такие исследования неизбежно несут. К сожалению, значительная доля (а по некоторым критериям — большинство) исследований на животных такой проверки не выдерживает.

Команда ученых под руководством невролога и специалиста по нейронаукам Малколма Маклауда в 2015 году опубликовала своеобразный обзор по исследованиям на животных. В нем отмечалось, сообщали ли авторы работ о следовании базовым принципам, касающимся дизайна исследования, — методикам, необходимым для того, чтобы делать надежные выводы по результатам такого рода экспериментов[36]. Первый принцип, который они проверяли, мы обсуждали выше: это рандомизация. Всякий, кто хоть что-то понимает в постановках экспериментов, знает, насколько рандомизация

важна, однако же в ходе анализа Маклауда обнаружилось, что лишь в 25 % релевантных статей сообщалось о рандомизации групп[37]. Не объясняется ли это тревожно низкое значение тем, что в исследованиях рандомизация применялась, но авторы об этом попросту не сообщили? Вряд ли: учитывая, какое решающее значение имеет рандомизация для качества данных, а следовательно, и для достоверности выводов (а в прагматическом отношении она здорово влияет на то, как статья воспринимается рецензентами), маловероятно, что исследователь, озаботившись рандомизацией групп, об этом не упомянет.

Другая методика, которую проверяли Маклауд и его коллеги, известна как "ослепление". В слепом исследовании ученый, собирающий данные, не знает ни проверяемой гипотезы, ни какая группа лечебная, а какая — контрольная[38]. Ученый обладает всей информацией, необходимой для проведения исследования, и не более того; и лишь когда все данные уже получены, он узнает, где какая группа. (В некоторых случаях ослепление также применяется к процессу анализа данных: ученый, проводящий статистические тесты, вплоть до их завершения не знает, какая группа была экспериментальной.) Ослепление должно проводиться на каждом этапе исследования, где только возможно: оно служит щитом, не позволяя осознаваемым или неосознаваемым предубеждениям испортить проведение эксперимента или анализ его результатов. Это хорошо известная концепция, которую преподают на каждом курсе по планированию экспериментов (по крайней мере, в отношении сбора данных, если не их анализа), и все же только в 30 % статей, посвященных исследованиям на животных, сообщалось о ее применении[39]. Причина может быть вот в чем: в некоторых своих предыдущих исследованиях, посвященных метаанализу, Маклауд обнаружил, что, когда ученые должным образом проводили рандомизацию и ослепление, в их экспериментах, как правило, обнаруживались гораздо меньшие эффекты[40].

Без рандомизации и ослепления даже исследования с громадным размером выборки могут вводить в заблуждение[41]. Но это вовсе не значит, что размер выборки не имеет значения. На самом деле это один из самых важных параметров, учитываемых при разработке эксперимента, и Маклауд с коллегами его тоже проверял. Сообщается ли в статье, как авторы решали, сколько животных включить в исследование? Об этом говорилось только в 0,7 % работ. Это удручает по двум причинам. Первую мы обсуждали раньше — *p*-хакинг. Не задавая предварительно размер выборки, исследователи оставляют себе возможность продолжать без конца собирать данные и анализировать их — снова и снова, снова и снова, пока не получится желаемое *p*-значение, меньшее 0,05. Вторая причина связана со смежным понятием, которое мы еще не обсуждали, — со *статистической мощностью*. Проще говоря, слишком много научных исследований слишком малы.

Представим себе идеальный препарат от головной боли, при любых условиях мгновенно ее снимающий. Нам не понадобились бы *p*-значения или статистические тесты, чтобы обнаружить этот сверхсильный эффект: мы замечали бы его всякий раз при сравнении хотя бы одного страдальца с головной болью, принимающего такие таблетки, с контрольным пациентом, получающим плацебо или менее эффективный препарат. Как если бы всякий мужчина на свете был выше абсолютно любой женщины, если вернуться к исследованию роста шотландцев из предыдущей главы. Конечно, на самом деле так не бывает: реальные статистические эффекты почти всегда меньше, и их труднее заметить. Реальная таблетка может уменьшить головную боль в среднем, скажем, на полбалла по шкале интенсивности мучений от 1 до 5. Невозможно было бы отделить такой небольшой эффект от случайного шума при сравнении двух человек, проводить подобное исследование было бы бесполезно. Даже если бы мы сравнивали две группы по десять участников, небольшой

эффект мог бы запросто затеряться в случайном шуме. Например, кто-то по невнимательности обвел бы неправильную цифру в анкете, или ударился головой перед опросом, что усилило боль, или, наоборот, облегчил свое состояние, бросив пить.

Но если бы мы включили в исследование гораздо больше людей (человек пятьсот, принимающих таблетки, и пятьсот — плацебо), скромный эффект от таблеток было бы куда легче отделить от случайных отклонений. Так получилось бы потому, что эффект от лекарства проявлялся бы *систематически*: наш сигнал менялся бы в одну и ту же сторону у достаточно большого количества людей, принимающих препарат. А вот шум оставался бы *случайным*: у людей из любой группы по причинам, никак не связанным с тем, принимали ли они таблетки или плацебо, боль иногда становилась бы то слабее, то сильнее. Поскольку число участников велико, эти случайные отклонения компенсировали бы друг друга, так что среднее значение в крупной выборке было бы ближе к “истинному” эффекту. Статистик сказал бы, что при большем размере выборки исследование имеет более высокую *статистическую мощность* — больше шансов обнаружить разницу между группами, если новый препарат действительно работает лучше, чем плацебо.

Как мы обсуждали в предыдущей главе, p-значение характеризует вероятность, что мы получим результаты, похожие на наши (или даже более впечатляющие), если на самом деле никакого искомого эффекта нет, поэтому обычно мы хотим, чтобы оно оказалось как можно меньшим (по крайней мере ниже стандартного порога, обычно устанавливаемого на уровне 0,05). С другой стороны, статистическая мощность характеризует вероятность, что мы увидим статистически значимый сигнал, когда он действительно *есть*, поэтому мы хотим, чтобы она была как можно *большей*. Меньшие эффекты (слабые сигналы) гораздо сложнее обнаружить, когда у вас мало данных, поэтому обычно чем

более тонкий эффект вы исследуете, тем крупнее вам требуется выборка.

Приведу конкретный пример. В 2013 году психолог Джозеф Симмонс со своими коллегами попросил онлайн-выборку участников ответить на ряд вопросов об их предпочтениях в таких областях, как еда и политика, а также собрал их основные анкетные данные (пол, возраст, рост и так далее)[42]. Затем он разделил выборку на различные группы (мужчины и женщины, либералы и консерваторы и тому подобное) и отметил, насколько сильно те различаются по целому набору переменных. Исходя из этих данных, Симмонс вычислил, сколько участников потребовалось бы, чтобы с уверенностью обнаружить данное различие, если не знать о его существовании заранее[43]. Например, оказалось, что можно надежно установить уже знакомую нам связь между ростом и полом — мужчины в среднем выше женщин — с помощью всего лишь шести мужчин и шести женщин из опроса; этот эффект, как мы знаем, велик и потому очевиден (наше исследование из предыдущей главы с участием двадцати человек имело, стало быть, высокую статистическую мощность). Еще один простой вопрос: склонны ли участники опроса более старшего возраста говорить, что они ближе к пенсионному возрасту? Так и есть, и Симмонс обнаружил, что для выявления этого факта потребовалось бы всего девять человек постарше и девять помоложе. Однако вот некоторые эффекты, для обнаружения которых понадобилось бы большее число участников:

- Любители острой пищи чаще уважают индийскую кухню (понадобилось бы двадцать шесть любителей острого и двадцать шесть человек, острое не любящих).
- Либералы считают социальную справедливость более важной, чем консерваторы (понадобилось бы по тридцать четыре человека от каждого политического лагеря).
- Мужчины в среднем весят больше, чем женщины (понадобилось бы по сорок шесть представителей каждого пола).

Цель этого упражнения состояла в следующем: заставить ученых реалистично оценивать величину искомого эффекта в любом конкретном исследовании, а значит, и размер выборки, необходимый для того, чтобы результаты получились значимыми. Если размер выборки у вас не будет достаточным для надежной проверки, весят ли мужчины больше женщин, скорее всего, статистической мощности вашего исследования не хватит и на то, чтобы обнаружить специфический тонкий эффект, подразумеваемый теорией, которую вы сформулировали.

Проводить исследование с низкой статистической мощностью — это все равно что искать далекие галактики с помощью бинокля: даже если искомый объект точно где-то там, увидеть его у вас нет фактически никаких шансов. К сожалению, многие ученые, похоже, упускают это соображение из виду, особенно в избранной Маклаудом области исследований на животных. В обзоре 2013 года рассматривался целый ряд работ по нейронауке, в том числе те, где изучались, например, половые различия в способности мышей ориентироваться в лабиринтах[44]. Чтобы исследование имело достаточную статистическую мощность для обнаружения обычно ожидаемого эффекта пола в способности ориентироваться в лабиринтах, нужны были бы сто тридцать четыре мыши; иными словами, этот эффект гораздо более тонкий, чем "мужчины весят больше женщин". Но типичное исследование из рассмотренных авторами обзора включало всего двадцать две мыши. И это относится не только к мышам в лабиринтах: похоже, проблема характерна для большинства типов исследований по нейронауке[45]. Еще в масштабных обзорах обнаружилось, что недостаточная статистическая мощность — обычное дело для медицинских испытаний, биомедицинских исследований в целом, экономики, нейровизуализации, исследований в области сестринского дела, поведенческой экологии и — какой сюрприз! — психологии[46].

Если исследования в этих областях имеют столь малую статистическую мощность, почему же тогда во многих

из них все-таки обнаруживаются какие-то эффекты? Первая причина заключается в том, что они, возможно, подвергнуты *p*-хакингу: ученые не обнаружили эффекта при своем первоначальном анализе, поэтому подошли к полученным числам творчески[47]. Но даже и без *p*-хакинга в исследованиях с недостаточной статистической мощностью все равно время от времени обнаруживаются какие-нибудь эффекты, и происходит это по тревожной, хотя и немного мудреной, причине. Вспомните, как мы обсуждали *ошибку выборки*. Представьте, что средний эффект нашей таблетки от головной боли в популяции действительно составляет полбалла по шкале от 1 до 5. Иногда мы можем взять выборку, для которой эффект по случайности ниже среднего, и все будет выглядеть так, будто эффекта нет и вовсе. А иногда попадется выборка, где эффект выше среднего, поскольку так получилось, что в нее вошли только те люди, которым лекарство помогало сильно. В исследовании с низкой статистической мощностью мы можем получить положительный результат — значимое *p*-значение — *только* в том случае, если выборка демонстрирует необычно и обманчиво большой эффект.

Рискуя прозвучать тавтологично, сформулирую так: поскольку исследованиям с малой статистической мощностью хватает ее только для обнаружения больших эффектов, это единственные эффекты, которые там проявляются. Вот куда приводит нас логика. Если вы обнаружили эффект в исследовании с недостаточной статистической мощностью, *он, вероятно, преувеличен*[48]. Затем вступает в силу публикационное смещение: поскольку большие эффекты впечатляют, статью о них, по всей вероятности, в итоге опубликуют. Поэтому-то, когда читаешь научную литературу, кажется, что так много крошечных исследований сообщают о больших эффектах: как мы видели на воронкообразных диаграммах в предыдущей главе, в журналах часто упущены все мелкие исследования, которые из-за того, что в них не обнаружилось ничего "интересного", были отброшены.

Эта ситуация создает проблемы для последующих исследований. Ученые разбирают опубликованную литературу, чтобы понять, какой величины эффекта ожидать в собственных экспериментах. Если в первоначальном небольшом исследовании величина эффекта преувеличивается, ученые для последующего изучения станут использовать маленькие выборки, полагая, что статистической мощности будет достаточно. Однако заявленный эффект, если он вообще существует, в действительности, вероятно, гораздо меньше и поэтому неуловим в экспериментах с малыми выборками[49]. Таким образом, исследования с недостаточной статистической мощностью запускают цепную реакцию: раз за разом впустую тратятся время, усилия и ресурсы в погоне за эффектом, который подобен гигантской тени, что отбрасывает мотылек, сидящий на лампочке.

Использование малых выборок не было бы столь пагубно, живи мы в мире, где и правда на каждом шагу встречались бы большие эффекты. Но обычно большие эффекты связаны с весьма очевидными факторами — вроде разницы в росте между мужчинами и женщинами, как в нашем примере. А большинство эффектов куда менее очевидны. В одном исследовании, посвященном клиническим испытаниям, обнаружилось, что медицинский эффект обычно бывает от малого до среднего. Грубо говоря, если в испытании участвовало сто человек, принимающих лекарство, и сто — плацебо и двадцати участникам стало лучше на плацебо, то человек на шесть больше (то есть около двадцати шести) почувствуют улучшение на лекарстве[50]. Даже для хорошо зарекомендовавших себя препаратов, таких как нейролептики при шизофрении, бензодиазепины при бессоннице и кортикостероиды при астме, эффекты все равно по величине лишь умеренные: в этих трех случаях плюс восемнадцати пациентам (итого получится около тридцати восьми) из лечебной группы станет лучше[51]. В исследованиях по психологии средний эффект тоже весьма скромен, и наверняка похожая ситуация наблюдается во многих других областях[52].

Когда речь идет об изучении таких необычайно сложных систем, как тело или мозг либо экосистема, экономика или общество, ученым редко удается найти единственный фактор, который оказывал бы огромное влияние на какой-то другой. Наоборот, большинство интересующих нас психологических, социальных и даже медицинских явлений состоят из множества мелких эффектов, каждый из которых играет свою маленькую роль. Например, если экономисты захотят объяснить, почему разные люди в их выборке имеют разный доход, им нужно будет учитывать, где живут участники, их происхождение, способности, личные качества и образование, налоговую систему их страны и ее изменения со временем и еще целую кучу других факторов и событий, которые на всех этапах жизни участников могли подтолкнуть их судьбу в том или ином направлении. Тот факт, что малые эффекты встречаются гораздо чаще и в совокупности оказывают влияние гораздо более значительное, чем эффекты большие, делает исследования с недостаточной статистической мощностью, которые рисуют наш мир полным этих больших эффектов, еще сильнее вводящими в заблуждение.

Один из самых конфузных примеров того, как исследования с низкой статистической мощностью сбивают ученых с пути, — это оживление вокруг так называемых генов-кандидатов. За последние лет десять генетики выучили болезненный урок об опасности исследований с низкой статистической мощностью. Давно было известно, в основном благодаря работам на близнецах, что рост и вес людей, результаты когнитивных тестов (IQ, коэффициент интеллекта), вероятность развития различных заболеваний и психических расстройств, а также многие другие характеристики зависят от генетических особенностей[53]. Однако только лет двадцать назад для генетиков стала широко доступна технология, с помощью которой можно было попытаться точно определить, *какие именно* участки ДНК с какими признаками связаны. Первые попытки были предприняты в исследованиях,

где ученые выделяли и изучали конкретный ген — “кандидат” — в надежде установить, могут ли изменения в нем вызывать изменение соответствующего признака.

Поначалу казалось, что попытки успешны. Появлялись все новые и новые исследования генов-кандидатов, дающие положительные результаты: вариации в гене *COMT* оказались связанными с показателями когнитивных тестов, вариации в гене *5-HTTLPR* — с депрессией, а в гене *BDNF* — с шизофренией, и это лишь несколько хорошо изученных примеров. Накапливались сотни и сотни исследований[54]. Обнаруживаемые эффекты часто впечатляли: так, в одной статье из престижного журнала *Nature Neuroscience* в 2003 году утверждалось, что способность запоминать у людей с одним конкретным вариантом гена *5-HT2A* на 21 % ниже[55]. С такими большими эффектами мы уверенно продвигались к пониманию генетических основ важных признаков. Еще генетики начали прояснять биологические “пути” между генами и признаками: например, было установлено, что ген *5-HTTLPR* связан с депрессией за счет того, что делает миндалину (область мозга, связанную с эмоциями) более реактивной, когда ее владельцу угрожает опасность[56].

В пору моего студенчества, с 2005 по 2009 год, исследования генов-кандидатов были предметом бурных обсуждений. К тому времени, как я получил ученую степень в начале 2014-го, они были уже почти полностью дискредитированы. Как так случилось? Основным фактором послужило то, что технология усовершенствовалась и изучать генотипы людей стало гораздо дешевле[57]. А значит, в генетических исследованиях теперь возможно было использовать куда большие выборки — размером во много тысяч или десятков тысяч образцов. К тому же генетики начали применять другой подход: теперь вместо только одного или горстки генов-кандидатов они одновременно изучали многие тысячи участков ДНК, различающихся у разных людей, проверяя, какие из них теснее всего связаны с рассматриваемыми при-

знаками. Такой подход называется полногеномным поиском ассоциаций (GWAS, *Genome-Wide Association Study*), и подобные исследования имели куда большую статистическую мощность, а следовательно, позволяли найти генетические варианты, оказывающие гораздо меньшее влияние на признаки, в дополнение к существенному влиянию уже хорошо известных генов-кандидатов[58].

Вот только полногеномный поиск ассоциаций не выявил для прежних генов-кандидатов больших эффектов[59]. И негде было их дальше искать: будь они реальны, выделялись бы. Вместо этого новейшие исследования с высокой статистической мощностью по поиску полногеномных ассоциаций продемонстрировали, что сложные признаки у человека связаны, как правило (за несколькими крайне редкими исключениями), со многими тысячами генетических вариантов, и каждый оказывает, похоже, лишь крошечный эффект[60]. Неоткуда было взяться большим эффектам для единичных генов, что шло вразрез с результатами всех предыдущих прославленных работ о генах-кандидатах. С тех пор усилия, направленные на повторение исследований генов-кандидатов при высокой статистической мощности, дали ровнехонькие отрицательные результаты для показателей тестов на коэффициент интеллекта, для депрессии и для шизофрении[61].

По прошествии времени кажется, что вся литература о генах-кандидатах — какой-то сюрреалистический опыт: ученые выстраивали величественное здание детальных исследований на шатком фундаменте — на основаниях, как мы теперь знаем, совершенно неверных. Вот как сказал об этом Скотт Александер в блоге *Slate Star Codex*: “Это не просто путешественник, вернувшийся с Востока и утверждающий, что там водятся единороги. Это путешественник, который описывает жизненный цикл единорогов, их рацион, все различные подвиды этих животных, какие куски единорожьего мяса самые вкусные и дает детальнейший отчет о схватке единорогов и снежного человека”[62].

Вся эта печальная история — хрестоматийный пример, предупреждающий об опасностях низкой статистической мощности. Исходные исследования генов-кандидатов, будучи мелкомасштабными, могли выявить лишь большие эффекты, а значит, о них и сообщали. Теперь понятно, что те большие эффекты были резко выделяющимися, нелепыми случайностями из-за ошибки выборки. Ожидалось, что в последующих работах тоже обнаружатся большие эффекты, поэтому размер выборок оставался довольно скромным. Так исследования выезжали на случайных результатах — и выстроилась цепь недостоверных открытий, ставших мейнстримом, золотым научным стандартом в этой области. Разумеется, какие-то отрицательные результаты проскакивали, а некоторые специалисты по метаанализу били тревогу по поводу низкой статистической мощности[63]. Но большинство исследователей генов-кандидатов упорствовали. Знай эти генетики историю своей дисциплины, они крайне настороженно относились бы к генам "большого эффекта": Рональд Фишер, статистик, популяризовавший *p*-значение и идею "статистической значимости", понял, что сложные признаки должны быть *полигенны*, то есть зависеть от многих тысяч генов "малого эффекта", еще в 1918 году[64].

К счастью для генетиков, технологические усовершенствования, благодаря которым цена генотипирования снизилась, привели к тому, что их идеи насчет генов-кандидатов в итоге были проверены в исследованиях по полногеномному поиску ассоциаций с надлежащей статистической мощностью — и не осталось сомнений, на правильном ли пути они были (нет, не на правильном). С тех пор генетики перешли к рутинному использованию больших выборок, и, хотя все еще существует несколько оплотов веры в гены-кандидаты, такой тип исследований почти вымер[65]. Но представьте, сколько других научных областей еще не столкнулись с подобным главным испытанием. Целые пласты научной литературы — базирующиеся на малых исследованиях

с неправдоподобно большими эффектами — вполне могут оказаться столь же ошибочными и наполненными миражами, как и исследования генов-кандидатов.

Справедливо ли с моей стороны называть ученых, которые проводят исследования с недостаточной статистической мощностью, недобросовестными или даже некомпетентными? Специалисты по нейронауке, например, могут возразить, что их исследования очень дороги, учитывая, что им часто приходится платить за лабораторных животных и их содержание или использовать непомерно дорогое оборудование вроде МРТ-сканеров мозга. При таких расходах небольшие исследования — это все, что они вправе себе позволить. Более того, многие исследования проводятся аспирантами и молодыми учеными, недавно получившими ученую степень, у которых не так много грантовых денег, если таковые вообще имеются. Когда я делал замечания по поводу низкой статистической мощности на научных семинарах, часто звучали ответы в таком духе: “Моим студентам нужно публиковать статьи, чтобы быть конкурентоспособными на рынке труда, и они не могут позволить себе проводить крупномасштабные исследования. Им приходится довольствоваться тем, что у них есть”. Это ярчайший образчик того, как ученых с благими намерениями систематически поощряют — кто-то сказал бы “заставляют” — идти на компромиссы, которые в конечном счете делают их работу ненаучной.

Как бы там ни было, объяснение, *почему* проводятся исследования с низкой статистической мощностью, их не оправдывает. Мы вернемся к вопросу о том, кто (или что) виноват в такой недобросовестности, позже. А пока нам нужно признать, что, когда научное сообщество дало свое коллективное согласие на проведение таких исследований, оно пренебрегло одной из своих основных обязанностей — или даже отреклось от нее. Тенденция таких исследований вводить в заблуждение означает, что они активно *отнимают* у нас знание: зачастую было бы лучше, если бы их вообще никогда не про-

водили. Ученые, сознательно занимающиеся исследованиями с низкой статистической мощностью, и рецензенты с редакторами, которые допускают крошечные исследования к публикации, отравляют научную литературу медленно действующим ядом, ослабляя доказательства, необходимые для прогресса. А в случае с работой на животных крайне трудно оправдать “эвтаназию” всех этих жизней, если проводимые исследования имеют настолько маленькую статистическую мощность, что никогда и не было ни единого шанса ответить с их помощью на соответствующие научные вопросы[66].

В общественных опросах ученых, как правило, называют высококомпетентными[67]. В ошибках, описанных в этой главе, угнетает больше всего то, что подавляющее большинство ученых действительно знают, как надо. Они знают, что надо тщательно проверять, нет ли опечаток и других оплошностей, знают, как важны рандомизация и ослепление, и знают (уже с 1950-х годов), что контаминация клеточных линий — серьезнейшая проблема в таких областях, как онкология. Они знают еще из университетских лекций по статистике, что статистическая мощность — фактор исключительной важности, особенно в мире малых эффектов. Крошечные выборки просто не в состоянии дать осмысленный ответ на большинство научных вопросов.

И все же недобросовестность — со стороны ученых, рецензентов и редакторов — приводит к тому, что статьи с подобными вопиющими ошибками появляются в научной литературе с удручающей регулярностью. Как и в случае с предвзятостью, мы никогда не сумеем избежать *всех* ошибок: случайные опечатки или другие огрехи неизбежны в любой сложной человеческой деятельности. Но из-за уникального положения науки в обществе — она раскрывает тайны, выступает арбитром в спорах и указывает в направлении объективности — ставки тревожно высоки, общество относится

к науке чрезвычайно серьезно. В ответ ученые должны предъявлять к себе гораздо более высокие требования.

Чем объясняется парадокс ученых, знающих, как надо, и работающих в сообществе, смысл существования которого заключается в организованном скептицизме, — и все равно оказывающихся в окружении ошибок, которых можно было избежать? Как вышло, что ученых, которых так часто стереотипно представляют сверхаккуратными и педантично относящимися к цифрам и данным, столько раз ловили на невнимательности? Мы обсудим возможные объяснения в третьей части книги. Однако прежде нам необходимо диагностировать еще один недуг науки. Быть аккуратным и добросовестным — значит в том числе сопротивляться выходу за пределы имеющихся данных, обуздывать себя при описании того, какие выводы можно сделать из каждого отдельно взятого исследования и каково его значение для мира. В таком случае научная культура, казалось бы, должна быть культурой в первую очередь сдержанности и скромности. К сожалению, это не так.

Глава 6
Хайп

> Следующая история о контакте с инопланетянами правдива. И под правдивостью я имею в виду лживость. Все они лживы. Но это ложь увлекательная. В конце-то концов, не это ли настоящая правда? Ответ — нет.
>
> Леонард Нимой, *"Симпсоны"*[1]

В 2010 году журнал *Science* опубликовал статью исследователей из НАСА, в которой описывался штамм загадочных бактерий, обитающих в калифорнийском озере Моно[2]. Это странное место: мало того, что вода там сильнощелочная и раза в три солонее, чем в океане, так над поверхностью озера еще возвышаются причудливых форм, похожие на сталагмиты башни из известняка, которые выглядят так, будто попали туда прямиком со съемочной площадки научно-фантастического фильма[3]. Подходящий антураж для ученых, жаждущих исследовать некоторые из самых экзотических видов в мире и посмотреть, какие уроки те могут преподать нам о химии жизни на других планетах.

Согласно результатам исследования, проведенного под руководством микробиолога Фелисы Волф-Саймон, адаптация бактерий к экстремальным условиям в озере привела к двум уникальным явлениям. Во-первых, они росли и размножались в среде, которая почти не содержала фосфора — а он считался одним из элементов, необходимых для всех форм жизни. Во-вторых, и это даже поразительнее, причина процветания была в том, что бактерии, названные Волф-Саймон GFAJ-1, изменили саму основу собственной ДНК, заменив фосфор на элемент, имеющийся в озере Моно в избытке, — на мышьяк. Конечно же, хорошо известно, что мышьяк — яд, и это сделало результат вдвойне удивительным: не только не-

слыханно было, чтобы другой элемент заменил в молекуле ДНК фосфор, но и мышьяк, особенно в таких высоких концентрациях, обычно токсичен. Для GFAJ-1 это очевидно не так: для них мышьяк (если точнее, *арсенат*, одно из соединений мышьяка, часто встречающееся в природе) играл противоположную роль, поддерживая жизнь бактерий[4]. Если бы результаты подтвердились, они не просто исполнили бы желание автора "дайте Фелисе работу", что, кстати, и зашифровано в названии бактерий (GFAJ — *Give Felisa a Job*)[5]. Они изменили бы наше представление о жизни как таковой. Волф-Саймон, безусловно, осознавала значение своего открытия: на пресс-конференции при объявлении результатов она описала, как они с коллегами "приоткрыли дверь к возможным формам жизни в других уголках Вселенной"[6].

Пора рвать учебники? Не торопитесь. Другие исследователи с самого начала относились к "мышьяковой жизни" скептически[7]. Микробиолог Розмари Редфилд из Университета Британской Колумбии заметила в исследовании недостатки и описала их в серии подробных сообщений в своем блоге[8]. Волф-Саймон критику просто проигнорировала. "Мы не собираемся участвовать в такого рода дискуссии, — сказала она журналисту. — Любое обсуждение должно пройти рецензирование так же, как и наша статья"[9]. Выглядело это абсурдно, ведь НАСА явно стремилось вызвать широкий общественный интерес, когда объявляло о публикации статьи. За несколько дней до того они сообщили всему миру в интригующем пресс-релизе, что новое открытие "повлияет на поиск доказательств существования внеземной жизни"[10]. Это вызвало шквал догадок; автор одного из популярных блогов тут же предположил, что признаки жизни, наверное, обнаружены на крупнейшем спутнике Сатурна — Титане[11]. И хотя обнародованные результаты исследования оказались не столь волнующими, НАСА все равно удалось представить их в напыщенных выражениях. "Определение жизни только что расширилось, — восторгался один из руководителей НАСА сразу после по-

явления статьи. — Продолжая прилагать усилия по поиску признаков жизни в Солнечной системе, мы должны… рассматривать жизнь такой, какой мы ее еще не знаем”[12].

Тем не менее довольно скоро желание Волф-Саймон исполнилось: обсуждение перешло на страницы журналов. То был достаточно редкий случай, когда издание публиковало суровую критику одной из своих предыдущих статей (казалось бы, это пример научного процесса, работающего именно так, как ему и полагается): в *Science* вышло не менее восьми “технических замечаний” к исследованию, в том числе один за авторством Редфилд, вместе с дерзким ответом Волф-Саймон и ее коллег[13]. Затем Редфилд и ее команда подвергли утверждения о мышьяковых бактериях строгой проверке[14]. Одно из главных заявлений Волф-Саймон заключалось в том, что, хотя бактерии в среде без фосфата (распространенного в природе соединения фосфора, химически сходного с арсенатом) и без арсената не размножаются, при добавлении последнего они это делают. Редфилд и ее коллеги попытались воспроизвести этот результат в своей лаборатории, однако не преуспели: бактерии GFAJ-1 вообще отказывались расти, пока в среду не добавят фосфат. А что касается ДНК этих бактерий, Редфилд и ее коллеги обнаружили в ней лишь минимальное количество арсената, после того как промыли образцы в воде. Самое подходящее объяснение было банальным: обычная контаминация. Похоже, арсенат, которым Волф-Саймон кормила бактерии, был загрязнен фосфатом — в количестве достаточном, чтобы обеспечить некоторый рост. А образцы ДНК у Волф-Саймон могли быть загрязнены, наоборот, арсенатом из среды, похожей на воды озера Моно, которую она создала в своей лаборатории. Независимая попытка воспроизвести исследование в Швейцарской высшей технической школе Цюриха дала результаты, аналогичные результатам Редфилд, что забило последний эмпирический гвоздь в крышку гроба мышьяковой жизни[15]. Жизнь все еще оставалась такой, какой мы ее и знали.

В эпизоде с мышьяковой жизнью есть чему порадоваться: необычайное утверждение было немедленно подвергнуто научным сообществом строгой проверке. Наука именно для того и существует, чтобы заставлять нас пересматривать свои мнения, и история с бактериями GFAJ-1 во многих отношениях послужила наглядным примером, как этот процесс должен идти правильно. И все же последствия были суровыми: после авантюры с мышьяком Волф-Саймон опубликовала еще только одну статью, а потом с исследовательской деятельности переключилась на преподавание. Между тем НАСА — пожалуй, самая известная в мире научная организация, внушающая в равной степени трепет и восхищение, — серьезно подорвало доверие к любым своим будущим пресс-релизам. Проблема, конечно же, была в том, что они перестарались с шумихой вокруг результатов исследования, загнав его авторов в оборонительный угол и в конечном счете обесценив их научные публикации.

Мы не должны исключать, хоть это и вызывает беспокойство, что идею раздуть такие предварительные результаты НАСА искренне считало удачной. Вероятнее, однако, что виной всему финансовое давление. Научные учреждения прилагают большие усилия, чтобы убедить своих спонсоров — в случае НАСА это правительство США, которое может урезать финансирование из федерального бюджета, — что занимаются они делом стоящим. Как выразился автор одной статьи, где анализировался случай с исследованием мышьяковой жизни, НАСА должно "непрерывно создавать ощущение своей значимости"[16]. Понятно, как это стремление, доведенное до крайности, может вынудить так сильно переусердствовать с пресс-релизом.

Похожие требования предъявляются, разумеется, и к ученым. Они зависят от грантов на поддержку своих исследований и работают в атмосфере поощрения ярких и парадных результатов, а не трудов ломовой лошади, которые увеличивают наше знание лишь по крупицам. В следующей главе мы

увидим, как такого рода объяснения могут помочь нам понять — хотя уж точно не оправдать — научные проблемы, обсуждаемые в этой книге. Однако сначала рассмотрим различные способы, какими ученые продвигают свои исследования, вводя всех в заблуждение. В этой главе мы будем говорить о том, что научные результаты отнюдь не просто искажаются популярной прессой, это сами ученые стали основным источником хайпа. Всякий раз, когда они раздувают или представляют в ложном свете один из своих результатов, они рискуют нанести урон нашей коллективной вере в науку. А иногда, когда шумиха достигает своей высшей точки, рискуют подорвать доверие к целым научным областям.

В центре истории о мышьяковой жизни был пресс-релиз НАСА. Многие люди не осознают, что пресс-релизы о научных результатах пишут не только пресс-секретари и пиар-агенты — ученые сами принимают в этом активное участие. Более того, иногда они составляют вообще весь пресс-релиз. Сценарий, когда невинный исследователь занимается своим делом, а средства массовой информации выхватывают у него из-под носа какой-то результат и раздувают из мухи слона, вовсе не норма[17]. Главная проблема с пресс-релизами состоит не в том, что они регулярно сообщают о потрясающих открытиях, которые оказываются ошибочными. Напротив, беда в том, что они раздувают результаты зачастую совершенно добротных научных статей, заставляя их выглядеть более важными, революционными или актуальными для жизни людей, чем они есть на самом деле. В исследовании 2014 года, проведенном учеными из Кардиффского университета, были проанализированы сотни пресс-релизов, посвященных научным изысканиям в области здравоохранения, и сопоставлены с самими описываемыми работами, а также с итоговыми новостными материалами[18]. Выяснилось, что в пресс-релизах использовались, как правило, три разновидности раскрутки.

Первая — *необоснованные советы.* Давались рекомендации, как читателям следует изменить свое поведение (например, им советовали выполнять определенные упражнения), что было чересчур упрощенной или топорной интерпретацией результатов исследования. Это касалось 40 % проверенных пресс-релизов. Другой разновидностью раскрутки был *межвидовой скачок*. Как мы видели ранее, многие доклинические медицинские испытания проводятся на животных, таких как крысы и мыши, — подобная практика известна как "трансляционные исследования" или "моделирование на животных"[19]. Идея в том, что базовые принципы работы, скажем, мозга, кишечника или сердца можно изучать на животной "модели", а затем — серьезными трудами — полученные результаты в итоге перенести на человека, что поможет разработать более эффективные методы лечения. Однако открытие чего-либо на мышах (или на клетках в лабораторной посуде, или в компьютерной модели) и применение на людях отделены друг от друга множеством этапов. Сначала должен пройти целый цикл разработки и валидации методик испытания, а затем уже сами исследования — кропотливый процесс, который может занимать десятки лет[20]. Подавляющее большинство результатов, полученных на мышах, около 90 %, в конечном счете на людей не переносятся[21].

Ученые, проводящие исследования на животных, конечно, прекрасно об этом знают. Тем не менее команда из Кардиффского университета обнаружила, что это не мешало им приукрашивать пресс-релизы, намекая или даже прямо утверждая, что их результаты, полученные на начальном этапе на животных, имеют важные последствия для человека. Это относилось к 36 % пресс-релизов. В свою очередь, в новостных материалах об исследованиях в области здравоохранения признание, что описываемая работа не проводилась на людях, зачастую прячут абзаце в восьмом или девятом. Психофизиолог Джеймс Хизерс создал забавный аккаунт в твиттере исключительно для того, чтобы повторять в нем вводящие в за-

блуждение заголовки новостей о трансляционных исследованиях, такие как "Ученые разработали вакцину, которая подавляет тягу к нездоровой пище" или "Симптомы, похожие на проявления болезни Альцгеймера, купируются веществами из моркови", снабжая их простым и четким дополнением: "...НА МЫШАХ"[22].

Третья разновидность раскрутки, обнаруженная командой из Кардиффского университета, пожалуй, самая постыдная. Все, а особенно ученые, должны знать, что *корреляция не означает причинно-следственной связи*[23]. Этому базовому принципу учат на каждом курсе по основам статистики, и он постоянно всплывает во всех общественных дискуссиях о науке, образовании, экономике и многом другом. Когда ученые анализируют данные *наблюдений*, которые собраны без какого-либо рандомизированного экспериментального вмешательства (скажем, при отслеживании, как у детей по мере взросления растет словарный запас), они, как правило, просто смотрят на корреляции. Тут нечего стыдиться: мы многое можем узнать о том, как разные вещи связаны друг с другом в нашем мире, и построение точной картины корреляций — непременное условие для понимания таких систем, как мозг или общество. Правда, нам нужно быть крайне осторожными при интерпретации этих корреляций. Если мы обнаружим, что употребление большего количества кофе коррелирует с более высоким IQ (кстати, так и есть), то нельзя будет сделать вывод, будто "кофе повышает IQ"[24]. Причинная стрелка может с таким же успехом указывать в противоположную сторону: вы умнее — и поэтому пьете больше кофе. Либо же и то и другое может быть обусловлено неким третьим фактором, например принадлежностью к более обеспеченной социально-экономической группе — что делает людей здоровее, а следовательно, повышает их IQ, а еще приводит к тому, что люди пьют больше кофе, поскольку в их социальных кругах это модно[25]. Незатейливые и набившие оскомину рассуждения, однако же в 33 % пресс-релизов, изученных учеными из Кар-

диффского университета, осторожность в отношении причинно-следственных связей отбрасывалась, и звучало все так, будто результаты наблюдений, показывающие корреляцию, получены в ходе рандомизированного эксперимента, в котором причинно-следственные связи как раз можно было бы установить[26].

Раскрутка в пресс-релизах была сопряжена с шумихой в новостях. Исследователи из Кардиффа обнаружили, что если пресс-релиз преувеличивал утверждение первым, то аналогичное преувеличение в средствах массовой информации встречалось в шесть с половиной раз чаще для утверждений о рекомендациях, в двадцать раз — о причинно-следственных связях и в баснословные пятьдесят шесть раз — о переносе на человека (когда пресс-релиз был написан аккуратнее, журналисты преувеличивали факты лишь незначительно). И хотя это исследование и само было всего лишь корреляционным, команда из Кардиффа в 2019 году продолжила его впечатляющим рандомизированным[27]. Они работали с университетскими пресс-службами: изменяли выбранные наугад пресс-релизы, добавляя в них необоснованные утверждения о причинно-следственных связях, и сравнивали реакцию на подправленные пресс-релизы и на те, что больше соответствовали фактам. Как менялись пресс-релизы, так менялись и заголовки: преувеличение порождало преувеличение. Другое исследование, проведенное в том же 2019 году, прояснило еще одну деталь: раздутые новости о здоровье действительно заставляли читателей больше верить в пользу лечения[28].

В эпоху “чурналистики”*, когда журналисты в вечном цейтноте нередко просто повторяют в своих статьях содержание пресс-релизов (сообщения о научных новостях частенько формулируются практически идентично пресс-релизам), ученые обладают огромной властью — и огромной ответственностью[29]. Ограничения, накладываемые системой рецензиро-

* От англ. *churnalism*.

вания, какими бы слабыми они ни были, вообще отсутствуют при взаимодействии со средствами массовой информации, и предубеждения ученых относительно важности собственных результатов могут проявляться беспрепятственно. К сожалению, как только пресс-релиз надувает пузырь шумихи, трудно заставить его лопнуть. В 2017 году было показано, что результаты лишь около 50% исследований в области здравоохранения, освещаемых в средствах массовой информации, в итоге подтверждаются в метаанализах (иначе говоря, 50% оказываются в целом воспроизводимыми). Результат и сам по себе достаточно скандальный, но его еще усугубляет тот факт, что эти метаанализы редко, если вообще когда-либо, освещаются в прессе[30]. А к тому времени ущерб может быть уже нанесен. Приношу свои извинения Джонатану Свифту: раскрученная наука летит на крыльях, а опровержения, прихрамывая, плетутся вослед*.

Пожалуй, та преходящая шумиха, что распространяется в новостных статьях, еще не самая вредная. А вот хайп, проникающий в книги, — совсем другое дело. Когда научно-популярные произведения ученых становятся бестселлерами, они создают идеи, которые остаются с нами надолго. В лучшем случае книги понятно излагают сложные научные результаты для широкого читателя без преувеличений и искажений, предлагая новые способы осмысления человека и мира. В худшем же — становятся раздольем для хайпа, “Диким Западом” далеко за пределами юрисдикции шерифов-рецензентов[31]. И опять грешит этим посильнее многих других моя собственная область — психология, которая легко прогибается под пользующиеся огромным успехом жанры самопомощи и жизненных советов.

* Оригинальная цитата: “Молва летит на крыльях, правда, прихрамывая, плетется вослед…” (Пер. М. Шерешевской. Цит. по: *Англия в памфлете*. М.: Прогресс, 1987.)

Особенно яркий пример — идея “подвижного образа мышления”. Обладать подвижным образом мышления — значит верить, что, если усердно работать, возможно улучшать свои умственные способности, чтобы они не оставались статичными на протяжении жизни. Чего вы хотите избежать, так это “застывшего образа мышления”, когда не верите, что можете развивать свои способности. Создательница такой концепции образа мышления Кэрол Двек, психолог из Стэнфордского университета, опубликовала сотни научных работ на эту тему, но именно благодаря ее книге “Образ мышления: новая психология успеха” идея действительно получила широкую известность. Двек заставляет понятие образа мышления звучать как потенциально судьбоносное. “Взгляды, которые вы для себя принимаете, глубоко влияют на то, как вы ведете жизнь”, — пишет Двек. И еще: “Когда на первый план выходит образ мышления, вы попадаете в новый мир”[32]. И правда, когда вы узнаете об образе мышления, “вы вдруг поймете великих — в науке и искусстве, в спорте и бизнесе — и тех, кто мог бы такими стать. Вы поймете вашего партнера, вашего начальника, ваших друзей, ваших детей. Вы увидите, как раскрыть свой потенциал — и потенциал ваших детей”[33].

Хотя книга Двек в основном состоит из поучительных историй, ее успех и влияние (как и самой идеи ментальных установок в целом) основаны на том, что автор — ученый. И не просто какой-то ученый, а известный во всем мире профессор престижного университета, делящийся, как Двек пишет в самом начале книги, результатами своих научных исследований. Идеи Двек вызвали сущее помешательство в образовании: в 2016 году опрос американских учителей показал, что 57 % прошли обучение по принципам установок на развитие, а 98 % согласились с тем, что использование идей о подвижном образе мышления в классе улучшило бы показатели учеников. Тысячи школ в Великобритании упоминают на сайтах о своей политике по вопросу установок на развитие[34].

А что говорят нам лучшие исследования по установкам на развитие? В 2018 году свыше трехсот работ, посвященных образу мышления, были подвергнуты метаанализу[35]. Авторы сосредоточились на изучении корреляции между подвижным образом мышления людей (оцененным с помощью опросников) и их успеваемостью в школе или университете, а также на экспериментах, в которых у студентов пытались вызвать большее стремление к развитию, чтобы улучшить их отметки. В обоих случаях эффект был реальным — но слабым. Корреляционная часть исследования показала, что образ мышления объясняет около 1 % различий в оценках. Попытка создать у студентов подвижный образ мышления, когда сравнивались экспериментальная группа, прошедшая тренинг по установкам на развитие, и контрольная без оного, дала результаты немногим лучше. Не будь от подвижного образа мышления вообще никакого эффекта, ожидалось бы стопроцентное совпадение между распределениями школьных оценок в этих двух группах (то есть распределения были бы идентичными). И хотя тренинг по установкам на развитие раздвинул распределения отметок, но лишь слегка: после него совпадение составило 96,8 %[36]. Это нельзя назвать большим влиянием.

Если распространить даже такое скромное преимущество на тысячи или миллионы студентов, то в совокупности можно принести приличную пользу[37]. Но не так Двек предпочла подать концепцию подвижного образа мышления, да и родители с учителями иначе не повалили бы покупать ее книгу. Вместо этого она преувеличила индивидуальные эффекты такого образа мышления, представив его почти откровением[38]. Риск подобного чрезмерного преувеличения в том, что учителя и политики начинают рассматривать такие идеи, как установки на развитие, как своего рода панацею для образования, хотя растрачиваемые на них время и ресурсы лучше было бы направить на решение клубка социальных, экономических и других проблем, из-за которых школьники не справляются с учебой. Реальность просто меркнет на фоне напы-

щенных заявлений в книге Двек, которые, кстати, противоречат интеллектуальной скромности, коей требует наука. Как мы видели в предыдущей главе, комплексные явления складываются из множества мелких эффектов — ученым положено это знать, они не должны продвигать идею, будто бы существует единственное "быстрое решение" для чего-либо столь сложного, как образование детей[39].

Будем к Двек милосердны: метаанализ провели больше чем через десять лет после выхода ее книги "Образ мышления" в 2006-м. Положим, тогда было неясно, чем эти результаты обернутся (впрочем, поэтому-то и важна только что упомянутая интеллектуальная скромность). У других ученых, переквалифицировавшихся в писателей, подобной отговорки нет. Психолог из Йельского университета Джон Барг был ведущим автором того исследования прайминга, когда мысли о пожилых заставляли людей идти медленнее, которое, как мы видели во второй главе, в 2012 году на большей выборке и в более строгих экспериментальных условиях воспроизвести не удалось[40]. А в 2017-м, спустя годы после провалившегося повтора и расцвета кризиса воспроизводимости в психологии в целом, Барг опубликовал бестселлер под названием "И глазом не моргнешь: неосознаваемые причины, по которым мы делаем то, что делаем"[41]. В книге обосновываются сильные подсознательные воздействия на поведение человека, однако автор не только не упоминает ни одну из больших проблем с воспроизводимостью в этой области, но еще и беспечно продолжает цитировать социально-психологические исследования — часто с крошечными выборками и спорными результатами, — чтобы сделать эффектные утверждения о человеческом поведении[42].

Во введении, например, Барг пишет, что неосознаваемые воздействия "могут даже повлиять на вашу будущую работу и зарплату, о которой вы сумеете договориться, — все зависит от того, какой напиток держит в руке ваш потенциальный работодатель или на каком стуле он сидит"[43]. Утвержде-

ние о стуле основывается на исследовании с пятьюдесятью четырьмя участниками, которое показало, что определенные типы людей склонны были занимать более расистскую позицию, когда сидели в офисном кресле Барга, что, судя по всему, "настраивало" их на ощущение собственной могущественности, по сравнению с ситуацией, когда они сидели в меньшем кресле по другую сторону стола[44]. Утверждение о напитке — а именно: после того как люди подержали в руках горячий напиток, они оценивали других как более приятных личностей, словно "согревшись" по отношению к ним, — родилось из исследования с участием сорока одного человека, которое не воспроизвелось на гораздо более крупных выборках[45]. Впрочем, даже если оставить за скобками вопросы воспроизводимости, заметьте, что ни одно из этих исследований никак не связано с "потенциальными работодателями". Барг беззаботно перенес выводы из контекста своей небольшой работы на студентах в условия, где эти выводы просто не проверялись, — классический случай, когда на основании ограниченных данных делаются неправомерные утверждения.

Книги, о которых мы говорили до сих пор, иллюстрируют, как авторы раздувают хилую доказательную базу[46]. Но вот пример суперпопулярной книги, чьи утверждения были раскритикованы как вопиюще неверное толкование научных данных. В 2017 году специалист по нейронауке Мэттью Уолкер, профессор Калифорнийского университета в Беркли, опубликовал книгу "Зачем мы спим", где рассказывается, что все мы должны спать по восемь часов в сутки, иначе будем страдать от ужасных последствий для здоровья (и не только)[47]. Как и другие обсуждавшиеся нами книги, она стала мировым бестселлером. Уолкер также с успехом выступил на конференции TED: его речь под названием "Сон — это ваша суперсила" собрала с десяток миллионов просмотров[48]. Ричард Смит, бывший редактор журнала *The BMJ*, назвал книгу "Зачем мы спим" "одной из тех редких книг, что меняют ваш взгляд на мир и должны преобразить общество и медицину"[49].

Заявления Уолкера — какие угодно, но только не усыпляющие. Уже в первой главе он пишет: “чем меньше вы спите, тем короче ваша жизнь” и “регулярный ночной сон продолжительностью менее шести или семи часов разрушает вашу иммунную систему, больше чем вдвое увеличивая риск заболевания раком”*[50]. Оба утверждения противоречат имеющимся данным. Писатель и исследователь Алексей Гузей в своей статье 2019 года попытался отследить источники многих утверждений Уолкера[51]. Во-первых, он обнаружил, что исследования на самом деле демонстрируют *U*-образную зависимость между продолжительностью сна и риском смертности: люди, которые каждую ночь спят дольше восьми часов, живут меньше, как и те, кто спит по пять часов или менее[52]. Во-вторых, с имеющимися данными не согласуется утверждение, что более короткий сон повышает риск заболеть раком из-за “разрушения” иммунной системы (кстати, это пример ошибочного вытягивания причинно-следственной связи из корреляционных данных): увеличение риска заболеть раком у тех, кто меньше спит, от силы слабое, а скорее всего, вообще нулевое[53]. Гузей также подверг критике целый ряд других утверждений, сделанных в книге, и заметил, что в одном случае Уолкер показал только кусок графика зависимости между сном и риском травм — опустив неудобную часть, демонстрирующую, что люди, спящие ночью по пять часов, реже получают травмы, чем те, кто спит по шесть[54].

Понятное дело, это вовсе не означает, что сон — или стремление спать побольше — неважен. И что другие части книги “Зачем мы спим” столь же неаккуратны. Однако мы видим, как преувеличения в науке иногда принимают катастрофическую форму, не ограничиваясь только лишь рассказами о самопомощи. Уолкер *мог* написать куда более аккуратную с научной точки зрения книгу, ни на шаг не выходя за рамки того, что показывают данные, но она вряд ли прода-

* Перевод В. Феоклистовой.

валась бы столь хорошо или удостоилась бы звания той, что "должна преобразить общество и медицину". А так книга может ввести читателей в заблуждение, нагнав ненужную тревогу по поводу продолжительности сна или заставив их терять время, уделяя сну больше часов, чем необходимо. Что же касается фактической точности, то уже сама плотность сбивающих с толку утверждений в книге, проданной в таком количестве экземпляров, должна была бы всех нас лишить сна.

Может, я упускаю из виду главное? Может, научно-популярным книгам как предприятию коммерческому и не нужно быть строго на сто процентов аккуратными, устойчивыми даже к придирчивой критике? Может, "легко усваиваемый" подход к описанию научных результатов, пусть и в немножко упрощенном виде, в целом целесообразен, поскольку популяризирует науку и вводит ее в жизнь людей? И разве мы не хотели бы, чтобы такие книги писали люди, которые хотя бы *кивают* в сторону доказательств? В подобном аргументе есть определенная ценность, но в долгосрочной перспективе это плохо. Отказ от фактов в пользу красивой истории чреват гонкой по нисходящей: научно-популярные книги будут издаваться все более оторванными от данных и неточными. Когда они неизбежно изобличаются или рекомендуемые в них изменения образа жизни не оправдывают надежд, порожденных хайпом, ущерб наносится репутации науки в целом. Книги, которые мы только что обсудили, написаны профессорами из Стэнфорда, Йеля и Беркли соответственно. Если даже ученые такого уровня не гнушаются преувеличением доказательств, то почему кто-то другой должен?

Простота научно-популярных книг также противоречит одному из фактов, которые к этому моменту повествования должны быть уже совершенно очевидны: наука сложна. Даже лучшие писатели силятся в увлекательной форме донести до читателя пилообразность научного прогресса — когда результаты часто противоречивы и неоднозначны, а самые надежные доказательства из имеющихся иногда внезапно под-

рываются новыми данными. Сглаживая эти тонкости и создавая впечатление, будто у сложных явлений есть простые, единичные причины и способы устранения проблем, подобные книги способствуют формированию ложного образа науки[55]. К сожалению, как мы увидим дальше, завышенные ожидания, подпитываемые популярной наукой, вероятно, даже начали влиять на саму научную практику.

Хорошие новости! Научные инновации развиваются! Во всяком случае, именно к такому выводу можно прийти, если принимать за чистую монету язык научных публикаций. В 2015 году было проанализировано, как часто определенные позитивно звучащие слова появлялись в аннотациях научных статей[56]. Аннотации — это резюме в самом начале статей, где ученые пытаются привлечь внимание читателей, и в условиях переполненного научного рынка им приходится все больше для этого стараться. В той работе был построен график: доля аннотаций, содержащих определенные слова, по годам начиная с 1974-го. Частота использования слов "инновационный", "перспективный" и "надежный" росла экспоненциально; эпитеты "уникальный" и "беспрецедентный" стали гораздо более распространенными (не парадоксально ли?); частота использования слова "благоприятный" неуклонно росла[57]. Эпитет "прорывной" почти не встречался примерно до 1999 года, когда по какой-то причине вдруг резко пошел вверх. В среднем частота использования позитивно звучащих слов в аннотациях за изученный период увеличилась почти в девять раз: в 1974 году всего 2 % аннотаций содержали такие самовосхваления, а в 2014-м — 17,5 %. Авторы иронично заключили: "Экстраполируя динамику частоты позитивно звучащих слов за последние сорок лет на будущее, мы предсказываем — слово «новый» появится в каждой [аннотации] к 2123 году"[58].

Сомнительно, что развитие научных инноваций действительно ускорилось наряду со всплеском гиперболических вы-

ражений в научных статьях[59]. Более вероятное объяснение заключается в том, что ученые стали пользоваться таким языком чаще, поскольку это отличный способ сделать свои результаты привлекательнее для читателей и, что, пожалуй, важнее, для рецензентов и редакторов знаменитых журналов. Самые манящие журналы сообщают на своих сайтах, что ждут статей, которые имеют “большое потенциальное влияние” *(Nature)*, являются “самыми значимыми в своей области” и “представляют новые и важные в широком контексте данные” *(Science)*, отличаются “необычайной значимостью” *(Cell)* или “исключительной важностью” *(Proceedings of the National Academy of Sciences)*[60]. Бросается в глаза отсутствие в этом списке каких-либо упоминаний строгости и воспроизводимости, зато воздадим должное ведущему медицинскому журналу мирового уровня *The New England Journal of Medicine* — на его сайте заявлено, что редакторы ждут статей, отличающихся “научной достоверностью, новизной и важностью”, именно в таком порядке[61].

Резкое увеличение числа позитивно звучащих фраз в научных журналах говорит о том, что хайп не ограничен лишь пресс-релизами и научно-популярными книгами — он пропитал сам способ написания учеными статей. В научном сообществе такой тип раскрутки часто именуется понятием из политики — *подачей под нужным углом*. В одной работе 2010 года авторы взяли репрезентативную выборку статей о рандомизированных медицинских исследованиях с отрицательными результатами (иными словами, о тех исследованиях, где не обнаружилось разницы между испытываемым видом лечения и плацебо) и проанализировали, часто ли информация в них подавалась под нужным углом, что определялось по наличию выражений, призванных отвлечь внимание читателя от отсутствия положительных результатов[62]. Попытками подчеркнуть достоинства исследуемого вида лечения, несмотря на то что испытания он провалил, грешили 68 % аннотаций и 61 % основных текстов статей. Как мини-

мум один случай подачи под нужным углом в каждом разделе (введение, методы, результаты, обсуждение) содержался в 20 % статей. А в 18 % подача под нужным углом начиналась сразу с названия.

Распространенная форма такой подачи под нужным углом — уклончивые формулировки, к которым ученые прибегают в случаях с *p*-значениями, не являющимися статистически значимыми. Вспомним, что мы обсуждали в четвертой главе: обычно, чтобы объявить эффект “статистически значимым”, требуется получить для него *p*-значение меньше 0,05. Статистик Мэттью Хэнкинс собрал коллекцию подлинных цитат из опубликованных статей, где *p*-значения упорно оставались выше этого порога, но авторы явно сильно хотели получить значимые результаты[63]:

- “тенденция в направлении к статистической значимости” (для результата с $p < 0{,}06$),
- “весьма значимый” ($p = 0{,}09$),
- “в значительной степени значимый” ($p = 0{,}065$),
- “едва ускользнувший от статистической значимости” ($p = 0{,}0789$),
- “колебался в районе статистической значимости” ($p = 0{,}061$),
- “очень близко подошел к границе статистической значимости” ($p = 0{,}051$),
- “не абсолютно значимый, но весьма вероятно” ($p > 0{,}05$).

Много научной литературы посвящено исследованиям подачи результатов под нужным углом, когда авторы отслеживают нарушения в собственной области. В 15 % статей по акушерству и гинекологии ученые представили свои не являющиеся статистически значимыми результаты так, будто те свидетельствуют о достоинствах лечения[64]. В 35 % публикаций о прогностических тестах при раке ученые использовали прием подачи под нужным углом, чтобы скрыть незначимость своих результатов[65]. В 47 % статей из ведущих жур-

налов об испытаниях средств от ожирения в той или иной степени применялась подача под нужным углом[66]. В 83 % публикаций, рассказывавших об испытаниях антидепрессантов и успокоительных средств, не обсуждались важные ограничения дизайна исследований[67]. В обзоре работ, посвященных нейровизуализации, авторы сделали вывод, что раскручивание корреляции до причинно-следственной связи "повсеместно"[68]. Некоторые формы подачи под нужным углом переходят в мошенничество или по крайней мере грубую некомпетентность: в обзоре 2009 года было показано, что из выборки исследований, опубликованных в китайских медицинских журналах и заявленных как рандомизированные контролируемые испытания, лишь в 7 % действительно использовалась рандомизация[69]. Даже метаанализы, как мы уже видели, небезупречны. В 2017 году в обзоре метаанализов, посвященных диагностическим тестам (скажем, анализам крови для выявления болезни Альцгеймера), было показано, что в 50 % из них делался вывод о действенности теста, хотя там фигурировали незначительные, статистически незначимые эффекты. В обзоре заключалось, что подача под нужным углом "могла вызвать в клинической практике неоправданный оптимизм относительно действенности тестов"[70]. Похоже, это еще один пример, как стремление ученых раздувать свои результаты вводит в заблуждение тех людей, кто больше всего на них полагается[71].

Подача результатов под нужным углом в итоге служит той же цели, что и преувеличения в пресс-релизах и книгах: ученые жаждут подчеркнуть яркость и важность своей работы, поскольку яркие и важные работы — это залог грантов, публикаций и восторгов. Проблема в том, что может возникнуть петля обратной связи: хайп взращивает ожидание простых и ясных историй со стороны финансирующих организаций, издателей и общества, а это означает, что ученые должны заниматься чрезмерным упрощением и украшательством своих работ еще больше, дабы поддерживать интерес и продолжать

получать финансирование. А уж собственно наука выбирается из этой петли обратной связи крайне нездоровой.

Теперь, когда мы увидели, как сопрягаются хайп в средствах массовой информации и раскрутка в научных статьях, обратимся к одной области, где такая петля обратной связи особенно сильна.

Всегда есть некая “развивающаяся” область, которая подвержена наибольшему хайпу. Как правило, несколько публикаций с легкими для понимания результатами в известных журналах подхватываются средствами массовой информации, общественный интерес возрастает, и ученые в этой области пускаются в некое безрассудство, подбрасывая в топку хайпа безответственные и преувеличенные утверждения. Затем в последующих экспериментах эти напыщенные заявления воспроизвести не удается, ажиотаж спадает, и наука возвращается на нормальные рельсы. К числу особо раскрученных областей относятся исследования стволовых клеток, генетика, эпигенетика, машинное обучение и нейровизуализация, а в последние несколько лет на звание “самой раскрученной” области претендуют исследования *микробиома* — бесчисленных миллионов микробов, населяющих наши тела[72].

Из-за хайпа на микробиом нацелено множество продуктов и методов лечения. Так называемые “пробиотики” — напитки или таблетки, восполняющие запасы “хороших” бактерий в кишечнике, — превратились в многомиллиардную индустрию[73]. Также растет интерес к терапии, известной как “фекальная трансплантация”[74]. Это когда образцы кала здорового донора, изобилующие различными микробами, трансплантируются пациенту — обычно в ходе колоноскопии, но иногда посредством проглатываемых капсул[75]. Хотя поначалу эта идея может показаться столь же неправдоподобной, сколь она неприятна, *существуют* убедительные доказательства эффективности по крайней мере одного типа фекальной

трансплантации — при рецидивирующих кишечных инфекциях, вызванных бактерией *Clostridium difficile* (или *C. diff*). В тяжелых случаях, когда антибиотики не в силах победить *C. diff*, но при этом уничтожили все нормальные бактерии, фекальная трансплантация позволяет пациенту получить дозу микробов от человека, чей кишечник находится в лучшей форме, что помогает выиграть в борьбе с патогеном[76].

Однако следует насторожиться, когда на микробиом ссылаются как на фактор, способствующий развитию заболеваний и состояний, которые не имеют очевидной связи с кишечником. Вот где заявления ученых и реальность начинают серьезно расходиться. При чтении научной литературы складывается впечатление, будто микробиом — причина и решение поистине поразительного множества психических и физических проблем. Например, появились исследования, в которых утверждается наличие связи между микробиомом и депрессией, тревожностью, а также шизофренией, а фекальная трансплантация предлагается для лечения, помимо прочего, сердечно-сосудистых заболеваний, ожирения, рака, болезни Альцгеймера, болезни Паркинсона и аутизма[77]. Идея якобы в том, что в процессе своей жизнедеятельности микробы в кишечнике могут производить вредные химические вещества, которые распространяются по всему организму, вызывая проблемы далеко за пределами кишечника[78].

Доказательства, лежащие в основе таких утверждений, обычно не слишком впечатляют. В случае с аутизмом разрыв между данными и шумихой особенно велик[79]. В 2019 году в статье, опубликованной в ведущем мировом журнале *Cell*, было рассказано об эксперименте, когда авторы взяли образцы фекалий у шестнадцати детей с диагностированным аутизмом или же без него и трансплантировали мышам[80]. Затем они разводили этих мышей в стерильных условиях, то есть потомство на протяжении всей жизни имело только микробы человека (расстройства аутистического спектра — это нарушение развития, поэтому любое влияние микробиома должно про-

являться с самого начала жизни). В зависимости от того, микробами какого типа доноров был заселен кишечник мышат, в заданиях, разработанных, чтобы вызывать у грызунов подобие симптомов аутизма, вели они себя по-разному. Например, когда их помещали в клетку с другой особью, мыши, которые получили микробы от детей с аутизмом, реже к ней подходили, что напоминает социальные нарушения при расстройствах аутистического спектра, а еще в клетке с опилками они проводили больше времени за закапыванием стеклянных шариков, что похоже на повторяющееся поведение, часто наблюдаемое у людей с аутизмом[81].

Вы, возможно, подумали, что связь между таким поведением у мышей и аутизмом у людей, мягко говоря, малоубедительна, а еще задались вопросом, может ли столь скромное число доноров считаться репрезентативным, если мы говорим о людях с аутизмом в целом[82]. Тем не менее авторы с радостью сделали впечатляющий вывод: "вмешательства, связанные с микробиотой, такие как пробиотики [или] трансплантация фекальной микробиоты… могут предложить своевременный и мягкий подход к решению пожизненных проблем при [расстройствах аутистического спектра]"[83]. Они выпустили пресс-релиз, в котором обсуждался "значительный" эффект от пересадки кала и снова утверждалось, что, судя по результатам их исследования, однажды пробиотики будут использовать для лечения симптомов аутизма[84]. Все это было чистейшим надуванием щек. Даже если не принимать во внимание ничтожный размер выборки в исследовании и довольно спорные допущения об эквивалентности поведения человека и мыши, в эксперименте вообще не проверялась способность пробиотиков или трансплантатов *облегчать* какие-либо предполагаемые симптомы "аутизма" грызунов, не говоря уже об их "эквиваленте" у людей.

Еще авторы, похоже, прибегли к подаче под нужным углом — кое о чем пытаясь умолчать. Их исследование включало и второй тест на социальное поведение, когда мышь могла вы-

брать, на что или кого потратить время — на другую особь или на “небольшой предмет”. Гипотеза заключалась в том, что мыши с микробиомом от людей, страдающих аутизмом, выберут вместо мыши-компаньона объект, однако никакой разницы обнаружено не было. Как отметил научный журналист Джон Брок в подробном критическом разборе этого исследования, авторы проскочили мимо неудобного результата, ограничившись единственным предложением, тогда как все результаты, оказавшиеся знаменательными, были представлены в статье на полноцветном графике[85].

Раздувание и подача под нужным углом такого крошечного, предварительного исследования — уже достаточно плохо, но все еще хуже. Когда биостатистик Томас Ламли получил данные, которыми располагали авторы, и попробовал воспроизвести их анализ, он обнаружил, что они неправильно провели свои статистические тесты. У них каждая мышь словно бы имела своего собственного донора человеческих фекалий, но на самом деле на сто мышей приходилось очень немного доноров[86]. Когда тесты были проведены надлежащим образом, сохранился только результат с закапыванием шариков, причем с почти пограничным *p*-значением в 0,03. Несмотря на столь серьезную критику, авторы, насколько мне известно, ответом не озаботились.

Не все исследования микробиома в корне статистически ущербны, как описанная работа с аутизмом и мышами, хотя многие из них столь же сомнительны с точки зрения безмерно раздутых выводов. В методологически сходном исследовании 2019 года утверждалось, что пересадка мышам микробиома пациентов с шизофренией может вызвать у грызунов симптомы психоза. В заключение было сказано, что полученные результаты “способны привести к новым стратегиям диагностики и лечения” шизофрении, что кажется более чем преждевременным[87]. Как бы то ни было, не исключено, что различия в микробиоме действительно играют свою роль в сложных причинах возникновения симптомов аутизма или шизофре-

нии либо каких-то других заболеваний из перечисленных выше у мышей или у людей[88]. Однако исследователям микробиома необходимо постепенно накапливать надежные результаты, а не вываливать средствам массовой информации каждое небольшое, возможно, подвергнутое *p*-хакингу исследование, в котором обнаруживается хоть какой-то эффект, утверждая, что это огромный научный прорыв. Вполне вероятно, что количество пресс-релизов связано с незрелостью научной области, когда со стороны средств массовой информации больше внимания уделяется областям с уймой “многообещающих” результатов, среди которых маловато хорошо воспроизводимых.

В последнее время в научном сообществе звучат призывы охладить хайповый пыл вокруг микробиома и связанных с ним методов лечения, а также улучшить качество исследований[89]. А пока грубо преувеличенные утверждения из статей и пресс-релизов создают видимость научного обоснования для множества бесполезных, вредных или попросту идиотских средств, имеющих отношение к микробиому: изготовленный с использованием микробов из кишечника элитных спортсменов пробиотический напиток, который якобы может улучшить ваши показатели, помешательство на “ирригации толстой кишки”, когда кишечник промывают водой, что сопряжено с жутковато звучащими рисками вроде “перфорации прямой кишки”, и компания, занимающаяся тестированием микробиома, которая позволяет вам напрямую узнать “национальность” вашего[90].

Повальные увлечения, подобные микробиомной мании, то усиливаются, то ослабевают, однако есть одна область исследований, которая постоянно порождает больше шумихи, вызывает больше интереса со стороны средств массовой информации и страдает от описываемых в этой книге недостатков больше, чем любая другая. Это, конечно же, пита-

ние. У медиа просто зверский аппетит на всякие псевдооткрытия: “Новая страшная наука, показывающая, что молоко для вас вредно”; “Английский завтрак — убийца: бекон повышает риск рака”; “Новое исследование демонстрирует, что яйца разобьют вам сердце”[91]. Учитывая огромный объем информации и количество противоречивых предписаний, как нам следует изменить свой рацион, неудивительно, что люди запутались, что им следует есть. После многих лет, в течение которых общественность кормили раздутыми результатами, она потеряла всякую уверенность и скептически относится к исследованиям в этой области[92].

Наука о питании — нутрициология — переживает, как и психология, свой собственный кризис воспроизводимости. Отчасти это может быть связано с мошенничеством: например, кардиолог Дипак Дас, опубликовавший десятки высокоцитируемых статей о пользе ресвератрола (это вещество, содержащееся в кожуре красного винограда, а следовательно, и в красном вине) для здоровья сердца, был уволен из Университета Коннектикута в 2012 году, когда выяснилось, что он подделал данные в девятнадцати исследованиях[93]. Отчасти же, наверное, виновата предвзятость: многие исследования финансируются пищевой промышленностью[94]. Более того, многие ученые сами придерживаются тех диет, которые изучают, что дает им личный стимул искать доказательства их пользы[95].

Наконец, отчасти проблема обусловлена многими другими видами предвзятостей и ошибок, с которыми мы теперь знакомы. Возьмем, к примеру, идею о том, что мы должны есть меньше насыщенных жиров и больше ненасыщенных, — краеугольный камень советов нутрициологов, повторяемый в бесчисленных рекомендациях по правильному питанию[96]. Однако же это не подтвердилось в метаанализе 2017 года, где сравнивались насыщенные жирные кислоты и полиненасыщенные по их связи с сердечно-сосудистыми заболеваниями и смертью[97]. Вероятно, причин тому три. Во-первых, налицо характерные признаки публикационного смещения — с од-

нобокими воронкообразными диаграммами, свидетельствующими о том, что исследования с малыми выборками и небольшими эффектами отправились в "картотечный ящик"[98]. Во-вторых, в одном из испытаний, заявленном как рандомизированное, была допущена ошибка, которая означала, что рандомизированным оно, похоже, не было[99]. В-третьих, многие исследования были спланированы неумело — помимо питания менялись и другие факторы, которые могли повлиять на результаты[100]. Сделанный в метаанализе вывод гласил, что убедительных доказательств пользы от замены насыщенных жиров ненасыщенными мало и что в предыдущих метаанализах — на результатах которых правительства нескольких стран основывали свои рекомендации по питанию — просто не были замечены все проблемы.

Можно с уверенностью утверждать, что значительная часть разрекламированных в средствах массовой информации исследований в области питания подвержена также и *p*-хакингу. Поскольку существует так много больших наборов данных с таким количеством значимых переменных — а участники исследований в области питания обычно заполняют так называемый опросник по частоте потребления различных пищевых продуктов, где указывают все, что они ели, скажем, за предшествующую неделю, — есть масса возможностей прочесывать данные в поисках всего, что окажется статистически значимым. Это частично объясняет бессистемную, противоречивую мешанину из существующих в области питания корреляционных исследований. В уже классической работе под заглавием "Все ли, что мы едим, связано с раком?" исследователи Джонатан Шонфельд и Джон Иоаннидис наугад выбрали из кулинарной книги пятьдесят ингредиентов, а затем проверили, не говорилось ли в научной литературе, что те влияют на риск развития рака[101]. Про сорок из них — говорилось, включая бекон, свинину, яйца, помидоры, хлеб, масло и чай (в общем-то, все ингредиенты того самого "английского завтрака — убийцы"). Некоторые

продукты риск вроде бы повышали, другие снижали, а некоторые в разных исследованиях оказывали различное воздействие. Числовые данные, как мы знаем, всегда зашумлены, поэтому логично ожидать, что литература будет выглядеть неупорядоченной. Тем не менее мы должны задаться вопросом, что вероятнее: что 80 % часто употребляемых продуктов из выборки действительно влияют, зачастую сильно, на риск развития рака или что низкие стандарты в этой области привели к публикации недоброкачественных, подверженных *p*-хакингу исследований, которые просто выехали на случайности, введя нас в заблуждение, будто многие из основных продуктов питания либо опасны, либо гарантируют долгую здоровую жизнь?[102]

Исследования в области питания часто не оправдывают ажиотажа, поскольку в значительной степени опираются на результаты наблюдений, а не экспериментов. Иными словами, во многих работах просто собираются данные о том, что люди едят, без каких-либо рандомизированных контролируемых испытаний. Взаимосвязь между продуктами, которые считаются здоровыми или нездоровыми в наблюдательных исследованиях, и продуктами, которые определяются как таковые в рандомизированных испытаниях, очень хлипкая — должно быть, с одним из двух типов исследований что-то не так[103]. Еще одна проблема в том, что отличия по здоровью могут быть обусловлены не продуктами, которые люди едят, а скорее культурными или социально-экономическими факторами, влияющими как на здоровье, так *и* на рацион. Кроме того, каждый компонент чьего-либо рациона, скорее всего, коррелирует с каким-то другим компонентом, что усложняет анализ эффектов: люди, потребляющие больше яиц, вероятно, также едят больше бекона и сосисок, а еще, наверное, многих других продуктов и питательных веществ, о которых вы и не подумали спросить. Существуют статистические методы, позволяющие “внести поправку” на такие помехи, но их сложно применить правильно, а основаны они на том,

что вы измеряете потребление каждого вида пищевых продуктов и питательных веществ, которые могут быть важны[104]. Точность подобных измерений сама по себе служит предметом серьезных споров, и идет удивительно ожесточенная дискуссия о том, как в наблюдательных исследованиях регистрируется потребление пищи. Некоторые исследователи называют опросники по частоте потребления различных пищевых продуктов “в корне неверными”, отчасти потому, что те полагаются на зачастую неточные воспоминания людей о съеденных продуктах[105]. Еще искажения происходят из-за предвзятости социальной желательности, когда участники не хотят сообщать, что за последние семь дней слопали, скажем, пять двойных чизбургеров[106].

Одно решение, предложенное для улучшения эпидемиологии питания, таково: ресурсы, расходуемые на курирование данных наблюдательных исследований, вместо этого направлять на проведение серии очень крупных, четких “мега-испытаний”, в которых, ко всеобщему удовлетворению, будут установлены факты об оптимальной диете[107]. Беда в том, что масштабные исследования, посвященные питанию, — какие угодно, но только не четкие[108]. Одно из крупнейших рандомизированных испытаний в эпидемиологии питания, в котором участвовало свыше семи тысяч человек, было обнародовано в издании *The New England Journal of Medicine* в 2013 году и посвящено было средиземноморской диете[109]. По сравнению с контрольными участниками, которым предписывалось придерживаться диеты с низким содержанием жиров, у тех, что придерживались средиземноморской диеты (больше белого мяса и морепродуктов; больше орехов и бобовых; больше оливкового масла), показатель, отражающий количество инсультов, инфарктов и смертей от сердечно-сосудистых заболеваний в течение следующих пяти лет, оказался значительно ниже. Испанские ученые, проводившие исследование, объявили, что после получения результатов и сами перешли на средиземноморскую диету.

Исследование, названное *Prevención con Dieta Mediterránea* ("Профилактика с помощью средиземноморской диеты"), сокращенно PREDIMED, привлекло все внимание средств массовой информации, какого только можно ожидать: "Средиземноморская диета показала, что предотвращает инфаркты и инсульты"[110], "Средиземноморская диета снижает риск сердечно-сосудистых заболеваний"[111]. Даже Комиссия по калифорнийскому ореху выпустила взволнованный пресс-релиз[112]. И кто может их винить? Как сообщили нам авторы PREDIMED, это было "масштабное исследование", которое "убедительно доказало"[113] преимущества средиземноморской диеты.

А затем появился Джон Карлайл. Возможно, вы помните Карлайла как бесстрашного расследователя в мире данных, который проверил тысячи "рандомизированных" контролируемых испытаний и обнаружил, что многие из них рандомизированными вовсе и не были. Исследование PREDIMED было одним из разоблаченных по методу Карлайла: числовые показатели просто не согласовывались с утверждением, что исследование было якобы надлежащим образом рандомизировано[114]. Авторы вернулись к своим данным и, конечно же, обнаружили несколько серьезных ошибок. Среди прочего выяснилось, что не каждому участнику случайным образом назначалась одна из двух диет: когда несколько участников приходили из одной семьи, им всем назначалась одна и та же диета. Между тем в одном из мест проведения испытаний исследователи случайно провели рандомизацию по клиникам, а не по участникам, так что всем, кто посещал одну и ту же клинику, назначали одинаковую диету. Поскольку на людей из одной семьи или одной клиники неизбежно влияют и другие общие факторы, становится невозможно приписать различия в результатах между ними и остальными участниками именно диете[115]. Эти и другие ошибки были допущены в отношении тысячи пятисот восьмидесяти восьми участников — 21 % от всей выборки.

Статья, на которую к тому моменту успели сослаться уже более трех тысяч раз, была отозвана и заменена исправленной версией в 2018 году[116]. Как водится, авторы утверждают, что подправленные результаты *еще убедительнее* доказывают преимущества средиземноморской диеты. Но есть причины оставаться настороже. Например, любопытно, что когда три типа нежелательных исходов, отслеживавшихся в исследовании, рассматривают по отдельности, то диета, похоже, влияет только на риск инсульта, но не на инфаркты или смертность[117]. Кроме того, исследование завершили раньше, чем планировалось изначально, поскольку эффект от диеты был настолько впечатляющим, — а это в области клинических испытаний считается сомнительной практикой[118]. А еще большую тревогу вызывает тот факт, что с использованием данных PREDIMED опубликовано уже больше двухсот пятидесяти статей, рассказывающих о других эффектах средиземноморской диеты, причем в некоторых работах, по-видимому, есть необъясненные числовые несоответствия[119]. Характерны ли и для них проблемы с рандомизацией, остается неясным.

Я поведал здесь и об исследовании PREDIMED не потому, что это уж особенно яркий пример хайпа, а потому, что это образчик наших лучших исследований в области, славящейся немыслимой раскруткой, и оно показывает нам, как даже олицетворение строгости могут подтачивать скрытые недостатки. Пожалуй, как и психология, эпидемиология питания *сложна*. Чрезвычайно тонкая физиологическая и психическая машинерия стоит за тем, как мы перерабатываем пищу и выбираем, что есть; данные наблюдений безумно зашумлены и несут на себе отпечаток несовершенств человеческой памяти; рандомизированные испытания могут споткнуться о сложности собственного управления. С учетом всех этих обстоятельств повышенный интерес средств массовой информации к исследованиям в области питания вызывает особое сожаление. Возможно, именно те научные вопросы, на которые общественность больше всего хочет получить ответы, —

что есть, как воспитывать детей, как разговаривать с потенциальными работодателями и так далее — относятся к областям наиболее мутным, сложным и внутренне противоречивым. А значит, ученые в этих областях тем более должны серьезнее относиться к тому, чтобы разумно доносить свои результаты до общественности.

Это, конечно, замечательно — отговаривать от хайпа и разглагольствовать о том, как все сложно, но ведь ученые до сих пор работают под бременем необходимости “выдавать” миру результаты. Общественность заслуживает того, чтобы ее держали в курсе последних научных открытий, на которые зачастую идут ее налоги. Есть ли способ информировать людей о научных результатах, избегая при этом злоупотреблений, с которыми мы столкнулись? Вот пример, показывающий, как все должно быть устроено.

Согласно научному консенсусу, ничто не может двигаться быстрее света. На этом основана специальная теория относительности Эйнштейна; это подтверждают все результаты, полученные пока в физике. Вот почему наблюдение, сделанное в 2011 году в эксперименте OPERA, было таким странным[120]. OPERA — это совместный проект в физике элементарных частиц, в рамках которого изучались субатомные частицы при прохождении через земную кору между лабораторией ЦЕРН в Женеве в Швейцарии и детектором в Гран-Сассо в Италии. Команда проекта, насчитывавшая примерно сто пятьдесят ученых из множества университетов, обнаружила, что нейтрино (частицы, похожие на электроны, но не имеющие заряда) добираются до места назначения слишком быстро — они достигали Италии на 60,7 наносекунды (миллиардных долей секунды) раньше того времени, что потребовалось бы для преодоления этого же расстояния свету[121].

Несмотря на лихорадочную перепроверку своих расчетов и оборудования, физики, работавшие в проекте OPERA, так

и не сумели обнаружить никаких ошибок. Результат со сверхсветовыми нейтрино казался реальным. Поскольку уже начали расползаться слухи, ученые решили сообщить об этом открытии миру. Они опубликовали препринт, в котором описали результат, а затем выпустили пресс-релиз[122]. Тут они могли бы воспользоваться стратегией НАСА с "мышьяковой жизнью" и живописать, как непостижимые новые результаты перевернули наше фундаментальное понимание Вселенной. Вместо этого они проявили осторожность. Никакого раздувания или подачи под нужным углом. Напротив, в их заявлении прямо подчеркивалась неопределенность: "Учитывая потенциальные далеко идущие последствия подобного результата, необходимы независимые измерения, прежде чем эффект можно будет либо опровергнуть, либо твердо установить. Вот почему сотрудничество OPERA решило обнародовать этот результат для более широкого и тщательного изучения"[123].

В сопроводительных заявлениях исследователи из проекта OPERA выражали растерянность, называя результаты "полной неожиданностью" и "очевидно невероятными". Затем они ждали, как средства массовой информации отнесутся к этой новости. Хотя всплыло несколько неудачных заголовков ("Ученые из ЦЕРН «обошли скорость света»", — сообщала газета *Daily Telegraph*; "Могут ли путешествия во времени стать реальностью?" — вопрошали в программе "Доброе утро, Америка" на канале *ABC News*), скептицизм все же сумел пробиться наружу[124]. В большинстве новостных статей приводились цитаты исследователей, говоривших о том, что результаты поистине странные, специальную теорию относительности еще никому не удавалось опровергнуть, требуется проверка в других экспериментах и так далее.

А потом они, конечно же, обнаружили ошибку. Плохо подсоединенный волоконно-оптический кабель приводил к тому, что время движения нейтрино занижалось. Когда кабель подключили правильно, время тут же стало соответствовать теории Эйнштейна[125]. Второй эксперимент, где делались

другие измерения, подтвердил: нейтрино не превышают скорость света. Увы, то был не конец: вся эта история стала позором для многих физиков из команды, некоторые из них казнились, что вообще не стоило писать о результатах исследования[126]. Несмотря на осторожность, с которой они отнеслись к освещению открытия в средствах массовой информации (и несмотря на то, что сотрудничество проголосовало за обнародование результатов), руководитель проекта и другое ответственное лицо незначительно проиграли в голосовании по вотуму недоверия и вскоре ушли в отставку[127].

Жаль, что им пришлось уйти в отставку, поскольку сотрудничество OPERA обращалось с неожиданным результатом почти образцово. Физики привлекли внимание к странному открытию, требующему проверки, избегая при этом шумихи и делая все необходимые оговорки, — тем самым преподав миру ценный урок научной неопределенности. За первоначальным шквалом новостных сообщений в дальнейшем последовало освещение развязки той истории[128]. Нужно ли быть большим фантазером, чтобы предположить, что, будь ученые из проекта OPERA психологами, а не физиками, они бы махнули рукой на перепроверку своих результатов и поспешили заключить договоры на написание книг под названиями “Преодолевая барьеры. Что скоростные нейтрино означают для вашей веры в себя” и “Сверхсвет. Новая наука субатомного успеха”?

Пожалуй, нечестно было бы приводить историю эксперимента OPERA в качестве примера с грамотно написанными пресс-релизами, поскольку событие в силу своей исключительной необычности — очевидное нарушение законов физики — привлекло бы достаточно внимания независимо от любой раскрутки. Но она все равно служит примером, как ученые с самого начала могут привнести в новостной цикл предусмотрительность, не поддаваясь чересчур уж распространенному порыву результаты раздувать. Представьте, если бы все пресс-релизы — и все научные статьи,

коли на то пошло, — сопровождались обязательными антиажиотажными заявлениями, напоминающими читателям, что результаты предварительные, и предостерегающими от того, чтобы заходить с ними слишком далеко.

Это, конечно, противоречило бы тому, как организована научная система. Хотя осторожность, сдержанность и скептицизм — основные достоинства науки, наша система поощряет прямо противоположное. Ученых вынуждает публиковать как можно больше статей и немыслимо раздувать результаты академическая система, которая стала для правильного развития науки препятствием. Эту систему и шаги, какими мы можем начать ее исправлять, мы и обсудим в заключительной части книги.

ЧАСТЬ III
Причины и пути исправления

Глава 7
Порочные стимулы

Если принцип которому ты следуешь привел тебя к такому концу какой прок от такого принципа?*

Кормак Маккарти "*Старикам тут не место*" *(2005)*[1]

Калифорнийские лесные пожары 2017 года охватили более миллиона акров, уничтожив тысячи зданий и убив сорок семь человек. Затем последовала самая дорогостоящая за всю историю штата ликвидация последствий стихийного бедствия, на которую ушло 1,3 миллиарда долларов[2]. Для организации работ привлекли инженерный корпус армии США, а он, в свою очередь, нанял местных подрядчиков для разбора огромных завалов, что оставил после себя огонь. Но инженерный корпус допустил критическую ошибку: он платил по тоннам. Чем тяжелее груз, тем больше денег получали подрядчики — и некоторые из них пользовались этим, доводя ситуацию до абсурда. Один свидетель сообщил, что видел, как рабочие "накачивали груз жидкой грязью". Другие подрядчики начали "слишком много копать": вместо того чтобы просто собирать всякие обломки, они рыли огромные новые ямы, иногда даже разрушая фундаменты жилых домов, и набивали свои грузовики землей и бетоном. В итоге правительству Калифорнии пришлось заплатить еще три с половиной миллиона долларов *после* ликвидации последствий, чтобы привлечь новых рабочих и засыпать все ямы, вырытые предыдущими подрядчиками.

То был классический случай *порочного стимула*. Инженерный корпус поощрял не очистку территорий, а просто на-

* Перевод В. Минушина. Сохранена пунктуация русскоязычного издания.

личие грузовика потяжелее, невольно создавая новые проблемы. Нетрудно вспомнить аналогичные примеры из других областей: стимулы для журналистов, поощряющие доходы, а не оригинальные репортажи, что приводит к появлению бездарных статей-приманок; стимулы для учителей, поощряющие школьные рейтинги, а не обучение, что порождает сомнительные отметки; стимулы для политиков, поощряющие краткосрочные результаты голосования, а не долгосрочные решения, из-за чего субсидируются отрасли промышленности, использующие ископаемое топливо[3]. В этой главе мы разберем стимулы, встроенные в сегодняшнюю научную практику, и зададимся вопросом, поощряют ли они объективность — или что-то совсем иное.

До сих пор мы обсуждали, как ученые фабрикуют данные, прячут в "картотечный ящик" и подвергают *p*-хакингу исследования, не проверяют на наличие ошибок и раздувают результаты. В итоге сложилась картина научной практики, в корне противоречащая идеалу науки, и мы подробно разбирали, как это произошло. Единственный кусочек пазла, пока остававшийся за кадром, — *почему*. Большинство ученых говорят, что выбрали свою карьеру из-за непреходящего интереса к природе, благодаря вдохновлявшему учителю или наставнику либо в силу желания принести пользу обществу[4]. И когда их прямо спрашивают, подавляющее большинство одобряет все четыре мертоновские нормы: универсализм, бескорыстность, коллективизм и организованный скептицизм[5]. Так почему же люди, ставшие учеными из-за любви к науке и ее принципам, в конечном счете так плохо по отношению к ней поступают?

Частично ответ на этот вопрос стал очевиден еще в предисловии, когда я рассказывал о своей попытке воспроизвести исследование паранормального и о мгновенном отказе из журнала: отрицательные результаты и исследования-повторения мало интересуют научные журналы, несмотря на свою исключительную важность для целостного представления ка-

кой-либо проблемы. Поскольку работы, сообщающие о положительных, ярких, новых, интересных результатах, поощряются куда больше других, это побуждает ученых создавать именно их в ущерб всему остальному. А чтобы убедить рецензентов и редакторов, что подаваемые на рассмотрение статьи действительно обладают всеми этими качествами, слишком многие ученые в итоге обходят или нарушают правила.

В этой главе мы заглянем еще дальше. И обнаружим, что система поощрения в науке порождает одержимость не только определенными видами работ, но и *самими публикациями*. Система побуждает ученых не заниматься наукой, а просто удовлетворять своим собственным извращенным требованиям. Порочные стимулы лежат в основе очень многих сомнительных практик, подрывающих наши исследования.

Иногда говорят, в основном серьезно, что последним истинным научным экспертом был Чарльз Дарвин. Он знал все, что можно было тогда знать о его области, естественной истории, в немалой степени благодаря масштабным экспедициям и сети собеседников-ученых, которых он, по его собственным словам, "донимал письмами"[6]. В наше время невозможно стать таким всезнающим экспертом, как Дарвин, ни в одной научной области. Ведь сейчас мы в научных трудах утопаем. Современный Дарвин должен был бы умудряться изучать порядка четырехсот тысяч новых статей по биологическим и биомедицинским наукам в год; а по всем научным дисциплинам за 2013 год, например, было опубликовано 2,4 миллиона новых работ[7]. При анализе данных за всю историю науки выяснилось, что рост числа статей ускоряется: в период с 1650 по 1750 год ежегодный прирост составлял 0,5 %, с 1750 по 1940 год — 2,4 %, а с тех пор — 8 %. Этот последний показатель означает, что вся научная литература удваивается в объеме каждые девять лет[8]. В некотором смысле это усовершенствование: коллективно люди сейчас знают о мире гораздо больше, чем в прошлые века. Но стоит

задуматься: отражает ли этот резкий скачок количества статей *только лишь* прирост наших знаний?

Есть причины думать иначе. Вероятно, самый известный пример стимула в науке, вызывающего тревогу, — это схема "деньги за публикации". С начала 1990-х годов китайские университеты проводят политику выплаты ученым (во всяком случае, в естественных науках) денежной премии за каждую статью, опубликованную в ведущем международном научном журнале. Подробности неясны (в одном из наиболее полных расследований этого вопроса отмечается, что многие договоренности по оплате держатся в тайне), но основная идея заключается в том, что денежное вознаграждение зависит от престижности журнала, в котором выходит статья, и существенно возрастает для топовых изданий[9]. Если ученый публикует статью в *Nature* или *Science*, он может рассчитывать, по крайней мере в некоторых китайских университетах, на вознаграждение, во много раз превышающее его годовую зарплату.

Похоже, в Китае эта политика наиболее распространена и потенциально выгодна, однако другие версии той же прямой схемы "деньги за публикации" зафиксированы в качестве государственной политики в Турции и Южной Корее, а также в некоторых университетах других стран, включая Катар, Саудовскую Аравию, Тайвань, Малайзию, Австралию, Италию и Великобританию[10]. Плата за публикации плохо сочетается с мертоновской нормой бескорыстности: ученые в своей работе должны быть движимы не финансовыми интересами[11].

Прямая схема "деньги за публикации" относится к числу наиболее грубых приемов, используемых университетами для того, чтобы побудить ученых публиковаться как можно чаще, однако еще исследователи находятся под более тонким — но все равно остро ощущаемым — финансовым давлением. На академическом рынке труда решения о найме и продвижении по службе основываются в немалой степени на том, сколько в вашем резюме значится статей и в каких журналах они опубликованы. Иметь за душой слишком мало ра-

бот и опубликовать их в безвестных изданиях — значит серьезно уменьшить свои шансы получить или сохранить работу. В американской академической системе решение, дать ли кандидату постоянную штатную должность (тогда старший преподаватель — это самая низкая ступень — становится доцентом, и с этого момента ему, по сути, гарантирована работа на всю жизнь), принимается в значительной степени с учетом тех же видов показателей производительности.

Почему же, спросите вы, университеты отдают предпочтение этому основанному на публикациях показателю, а не другим, которые могут лучше отражать качество исследований, — например, соответствует ли работа ученого таким стандартам, как рандомизация и ослепление или даже воспроизводимость? Ответ: они тоже находятся под финансовым давлением. Во многих странах, включая Великобританию, сами университеты оцениваются правительством по престижности статей, публикуемых их сотрудниками, и деньги налогоплательщиков распределяются соответственно такому рейтингу[12]. Все это и породило расхожее выражение "Публикуйся или погибни": беспрерывно клепай статьи, причем в самые представительные журналы, куда только можешь, — или просто не выживешь в мире суровой конкуренции современной академической науки[13].

И речь не только о статьях. Мы уже говорили, что обычно первый необходимый шаг при проведении научного исследования — получить грант, чтобы иметь возможность платить за оборудование, материалы, доступ к данным, выплачивать вознаграждение участникам и зарплату персоналу. Это означает, что ученые должны постоянно подавать заявки на гранты, чтобы поддерживать свои исследования. И опять-таки университеты испытывают такое же давление. Они берут часть всех грантов, какие только удается выиграть их ученым, и тратят на субсидирование преподавания, наем сотрудников, содержание зданий и так далее. По этой причине они сильно зависят от своих ученых, приносящих деньги. В одном исследовании,

проведенном в США, было подсчитано, что ученые в среднем тратят около 8 % своего общего рабочего времени и 19 % своего исследовательского времени на написание заявок на гранты — и это еще, мне кажется, довольно низкие оценки[14].

Необходимость постоянно искать финансирование не просто отнимает время, но еще приводит к огромному числу неудач и разочарований. Проблема усугубляется так называемым эффектом Матфея: при распределении научных грантов богатые становятся еще богаче. (Это отсылка к Евангелию от Матфея, глава 25, стих 29: "Ибо всякому имеющему дастся и приумножится, а у неимеющего отнимется и то, что имеет"*[15].) Есть веские доказательства, что так и происходит: в одном крупном исследовании было показано, что ученые, чьи самые первые заявки на гранты были оценены как *чуть превышающие* произвольный порог для получения финансирования, в последующие восемь лет получили примерно вдвое больше денег по другим грантам, чем ученые, чьи заявки были оценены как *чуть не дотягивающие* до порога, хотя качество первоначальных заявок не могло настолько уж отличаться[16]. Из-за такой атмосферы многие ученые разочарованно уходят из профессии, а те, кто остается, вынуждены прибегать к преувеличениям и раскрутке в своих заявках на гранты, чтобы конкурировать за деньги с хорошо финансируемой старой гвардией. Атмосфера токсичная, и нетрудно представить, как точность самой науки отходит на второй план[17].

Теперь оставим в стороне финансовые стимулы и не будем забывать о человеческой природе: люди — кто больше, кто меньше — обладают естественной склонностью интенсивно конкурировать за статус, стремиться к достижениям, улучшающим репутацию, и работать даже ради объективно бессмысленных целей — в данном случае большого количества публикаций и грантов. Таким образом, для тех из нас, кто более амбициозен и склонен к соперничеству, длинное

* Синодальный перевод.

резюме может само по себе служить наградой. Кем-то просто упоминание его имени в научной статье — любой — ощущается как достижение.

Так или иначе, поощрение ученых больше публиковаться свой эффект, похоже, дает. Мало того, что резко выросло число публикаций, еще и отбор на продуктивность среди ученых, кажется, становится сильнее. Во французском исследовании было показано, что у молодых эволюционных биологов, принятых на работу в 2013 году, публикаций почти вдвое больше, чем у тех, кто трудится с 2005-го, а это значит, что критерии приема на работу из года в год повышались[18]. Весьма уместно, что данные относятся именно к эволюционной биологии, поскольку конкуренция за рабочие места — это именно тот вид отбора, что дал павлину его хвост, а лосю рога. В конкуренции за ограниченные ресурсы (для павлина и лося это возможность спариваться; для ученых — гранты и должности) те, кто обладал более выраженными признаками, продолжали побеждать, пока не обрели абсурдно демонстративные черты — или нелепо длинные резюме. И эти ограниченные ресурсы становятся все скуднее: хотя количество присуждаемых ученых степеней возросло (отметим, что это еще одно следствие того, как университеты заботятся о своих доходах, ведь аспиранты и другие студенты тоже приносят огромные деньги), университеты не успевают увеличивать число своих рабочих мест, которые должны были бы занять эти новоиспеченные ученые со степенью[19].

Возможно, вы задаетесь вопросом, почему это ученые не должны погибать, если они не публикуются. Разве не логично университетам желать, чтобы их исследователи делали больше открытий, причем поинтереснее, и делились ими со всем миром, публикуя статьи в известных журналах? Разве это не подходящий и эффективный способ оценивать их успехи? И почему ученые *не должны* окупать себя, конкурируя за гранты, чтобы только лучшие идеи выигрывали деньги? Разве иначе мы не взрастим попросту нахлебников,

бездельничающих в университетах и не вносящих никакого вклада в приумножение наших знаний?

В идеальном мире подобное стимулирование продуктивности имело бы смысл. В журналах был бы такой контроль качества, а ученые обладали бы такой врожденной добросовестностью, что, несмотря на растущее число публикуемых работ, их качеству ничего бы не угрожало. В реальности же приходится чем-то жертвовать. Применительно к экспериментам по когнитивной психологии, где участники должны как можно быстрее нажать на кнопку тогда и только тогда, когда видят световую вспышку, исследователи говорят о “компромиссе между скоростью и точностью”. Когда испытуемые сосредоточены на скорости, страдает точность; если они сосредотачиваются на том, чтобы сделать все правильно, им приходится замедляться. (И кстати, это любопытный и четкий психологический эффект, который действительно воспроизводится почти во всех релевантных экспериментах[20].) С научными публикациями дело обстоит так же[21]. Время ограниченно. Принуждение ученых публиковать все больше и больше статей и приносить все больше и больше денег — наряду с выполнением всех остальных обязанностей, таких как преподавание, курирование чужих работ и решение административных задач, — почти неизбежно подразумевает, что на каждое исследование они будут тратить меньше времени. Принуждение рецензентов (которые, конечно же, и сами занятые ученые) рассматривать все больше и больше поданных в журналы статей означает, что большее число ошибочных, чрезмерно раздутых или даже мошеннических исследований пройдет через этот фильтр. Неудивительно, что в обоих случаях стандарты снижаются[22].

Доказательством того, что по крайней мере для некоторых ученых количество преобладает над качеством, служат хитроумные способы, найденные ими, чтобы использовать систему

в своих интересах. Вот пример из исследования, посвященного анализу китайской схемы “деньги за публикации”, на которое я уже ссылался выше:

> Профессор Гао из Хэйлунцзянского университета опубликовал двести семьдесят девять статей в одном журнале, *Acta Crystallographica Section E*, и получил более половины всех денежных вознаграждений, выплаченных этим университетом в период с 2004 до 2009 года. <… > Все исследования профессора Гао в эти пять лет сводились к тому, чтобы находить новые кристаллические структуры в своей лаборатории и всегда подавать результаты в один и тот же журнал, ведь так он мог достичь своей цели в виде денежной премии в короткий срок, тогда как, занимаясь долгосрочными исследовательскими проектами, он получал бы меньше премий[23].

В этом отношении профессор Гао — академический эквивалент группы *Status Quo*, которая добилась успеха, с 1970-х годов выдавая бесконечные незначительные вариации пары основных рок-песен. И это очень непохоже на Чарльза Дарвина. Гао занимался тем, что стали называть “нарезкой”: брал набор научных результатов, зачастую из одного исследования, которые можно было бы опубликовать вместе как одну работу, и разделял на более мелкие подстатьи, каждую из которых публиковал отдельно[24]. Это ближайший аналог ситуации, когда водители грузовиков, занимавшиеся ликвидацией последствий стихийного бедствия, загружали кузовы жидкой грязью, просто чтобы сделать их потяжелее: ваше резюме искусственно раздувается и создает впечатление, будто вы провели гораздо больше исследований, что, по меньшей мере в некоторых академических системах, максимизирует вашу прибыль. В общем, представьте, что я опубликовал каждую главу этой книги, а также предисловие и эпилог по отдельности и заявил, что написал десять книг, а затем за каждую из них получил отдельную плату.

Расскажу про один из самых возмутительных случаев нарезки, с которым я недавно столкнулся. Иногда генетики, изучающие гены, связанные с психическими расстройствами, используют полногеномный поиск ассоциаций (GWAS), который мы обсуждали в шестой главе. У человека двадцать три пары хромосом, несущих в себе генетический материал, и стандартный метод GWAS предполагает, что исследуются все пары сразу (отсюда и слово “полногеномный”), чтобы найти любые связи с рассматриваемым признаком. Однако те генетики, о которых идет речь, вместо того чтобы провести стандартный широкомасштабный анализ, провели его для *каждой пары хромосом отдельно*, то есть превратили единственную статью, какую обычно пишут по результатам подобного исследования, в потенциальные двадцать три отдельные публикации. На момент написания этих строк они успешно опубликовали шесть[25]. Хотя такая наглость даже забавна и, безусловно, благотворна для резюме авторов, а вероятно, и банковских счетов, для науки это балласт. Заинтересованным читателям в итоге придется прочитать почти две дюжины статей, чтобы найти информацию, которая должна содержаться в одной. Это лишнее беспокойство и трата времени редакторов и рецензентов, вынужденных проверять каждую отдельную статью. Кроме того, ученые совестливые, которые объединяют свои результаты в большие, полноценные статьи, в мире, где тоненькие публикации могут стать подспорьем на рынке труда, оказываются в невыгодном положении.

Нарезка сама по себе не означает, что наука, содержащаяся в каждом из кусочков, обязательно плохого качества (хотя тот факт, что исследователи готовы столь откровенно эксплуатировать публикационную систему, не говорит об их благонадежности). Тем не менее иногда нарезка может иметь более низкую цель, чем просто наращивание резюме. Есть подозрение, что в области клинических испытаний фармацевтические компании и другие исследователи лекарств расчетливо

используют нарезку, дабы воспользоваться тем обстоятельством, что читатели не уделяют должного внимания каждой публикации. Разделите свое исследование на несколько публикаций — и вы создадите впечатление, будто эффективность вашего препарата подтверждена надежнее, чем если бы по нему были опубликованы лишь одна-две работы. Это хитрый, но, вероятно, эффективный прием: занятые врачи, видя, что есть шесть статей, свидетельствующих в пользу одного препарата, и только одна — в пользу другого, с большей вероятностью пропишут пациенту первый, не всегда замечая, что шесть статей о нем — это нарезка отчета об одном и том же исследовании. Более того, поскольку разные врачи читают разные журналы, нарезка публикаций позволяет охватить более широкую аудиторию.

Выяснилось, что работы, посвященные антидепрессанту дулоксетину, подверглись тончайшей нарезке. Вот лишь один пример: исследователи дулоксетина опубликовали одну статью про различия в действии препарата, обусловленные этническими особенностями чернокожих и белокожих людей, а другую — о различиях в действии препарата между латиноамериканцами и белыми, хотя все данные были получены в ходе одних и тех же испытаний[26]. Казалось бы, нет никаких причин разделять эти два анализа, их нужно публиковать вместе в одной статье. Никаких причин, кроме того, что большее количество статей является частью “публикационной стратегии”, которую фармацевтические компании устанавливают для своих препаратов. Это маркетинг, не наука, а расплачиваются пациенты, чьими врачами сознательно манипулируют, заставляя выписывать лекарства, которые могут быть гораздо менее эффективны, чем они думают[27].

Еще одним свидетельством того, что количество ставит под удар качество, служит появление так называемых хищнических журналов. За последние пятнадцать лет резко выросло число сайтов, которые для неподготовленного человека выглядят настоящими научными изданиями, но не придержи-

ваются ни одного из обычных стандартов рецензирования и других правил, принятых в научных журналах[28]. Этими потемкинскими изданиями владеют недобросовестные предприниматели, которые пытаются эксплуатировать хорошо известное желание ученых иметь побольше публикаций. Они засыпают электронные почтовые ящики спамом, часто на ломаном английском, и в этих письмах упрашивают ученых подать к ним свою работу, хвастаясь тем, как быстро их журнал принимает статьи к публикации. Увы, многие неопытные, беспечные или отчаявшиеся исследователи попадаются и публикуются там (и конечно же, с них взимается плата якобы за "обработку" статьи), подрывая при этом свою репутацию: раз ученый опубликовал статью в поддельном журнале, значит, он либо легковерный, либо бесчестный[29].

Обычно хищнический журнал легко отличить от настоящего: сайты оформлены ужасно, статьи напечатаны плохо, а редакционная коллегия состоит из сотрудников таких университетов, о которых едва ли кто-то слышал, а иногда и из таких, что и вовсе не существуют. Зачастую, однако, грань между реальной и хищнической деятельностью размывается. Попытки составить списки поддельных изданий значительно усложняются невозможностью дать определение "хищническому" журналу, которое бы всех удовлетворило, угрозами юридического характера со стороны сомнительных издательств, а также скоростью появления все новых и новых поддельных журналов[30].

Худшие хищнические журналы опубликуют буквально все, даже самую очевидную утку. В 2014 году специалиста по компьютерным наукам Питера Вэмплю настолько достал непрекращающийся поток спама, рассылаемого хищническим журналом *International Journal of Advanced Computer Technology*, что он подал туда шуточную статью под названием "Удалите меня из вашего гребаного списка рассылки". Статья состояла из единственного предложения "Удалите меня из вашего гребаного списка рассылки", повторенного свыше восьмисот раз (еще там была полезная блок-схема с квадрати-

ками и стрелочками, отображающая основную мысль: *Get* → *me* → *off* → *Your* → *Fucking* → *Mail* → *ing* → *List*). Журнал оценил статью на отлично и принял к публикации[31].

Ни нарезка статей, ни публикации в хищнических журналах явно не противоречат никаким правилам, и определить, что именно считается "нарезкой", так же трудно, как и разделить все журналы на "хищнические" и легитимные. Но это не значит, что в погоне некоторых ученых за все большим числом публикаций не происходит откровенного нарушения правил — то есть мошенничества. Как мы уже видели, мошенничество в науке присутствует неотлучно, и оно влияет на процесс публикации точно так же, как и на процесс сбора данных и написания статьи. Например, мошенники пользуются тем удивительным, возможно, фактом, что ученые часто могут предлагать рецензентов для своей собственной статьи, когда подают ее в журнал. Редактор волен отправить статью предложенным рецензентам или же обратиться к выбранным им самим, но чаще предпочитает первый вариант: идея предлагать рецензентов, собственно, и заключается в том, чтобы избавить редакторов, у которых совершенно нет времени, от необходимости искать требуемых экспертов в соответствующей области. Система, что понятно, подвержена злоупотреблениям: авторы могут предлагать в рецензенты своих друзей или коллег, облегчая себе путь к публикации. Это довольно плохо, но, как всегда, мошенничество выводит все на новый уровень. Один редактор описал случай с биологом Мун Хён Ином (тринадцатый* номер в "списке лидеров" *Retraction Watch* с тридцатью пятью отозванными статьями):

> Он предлагал предпочтительных рецензентов… [и это] были он сам или его коллеги под фиктивными именами и аккаунтами. В некоторых случаях он указывал имена реальных людей (если бы вы стали искать их в *Google*, то увидели бы, что

* Уже шестнадцатый, конкуренция высока.

> они действительно существуют), но создавал для них учетные записи электронной почты, к которым он или его коллеги имели доступ и которые затем использовались для отправки замечаний от рецензентов. В других случаях имена и адреса электронной почты он просто выдумывал. Замечания этих рецензентов почти всегда были благосклонными, но все же содержали предложения по улучшению работы[32].

Редакторы заподозрили неладное, потому что рецензии часто приходили в течение суток. Дурацкий прокол Муна: настоящие ученые, которые, как известно, занятые и иногда запаздывают с рецензиями на недели или даже месяцы, никогда не проявлялись бы с такой упорной пунктуальностью. Мун далеко не одинок: фальшивые рецензии — классика жанра в базе данных *Retraction Watch*[33]. В 2016 году крупное научное издательство *Springer* впало в такое отчаяние из-за свирепствующего мошенничества с рецензиями в одном из своих журналов, *Tumor Biology*, что после отзыва ста семи нечестных статей из всего лишь четырехлетнего выпуска отказалось от издания этого журнала и продало его другой компании[34].

Слабым утешением здесь может послужить довольно удручающий факт, свидетельствующий об уменьшении ущерба, который наносят подобные недобросовестные публикационные практики. Вот он: огромное количество этих статей практически не привлекают внимания других ученых. В одной работе было показано, что за пять лет с момента публикации примерно на 12 % статей о медицинских исследованиях и на 30 % статей по естественным и социальным наукам ни разу никто не сослался[35]. Не исключено, что эти всеми заброшенные работы в конце концов будут кем-то процитированы, а может, в процессе анализа какие-то цитирования оказались пропущены[36]. Но хотя, наверное, хорошо, что низкокачественные продукты системы по максимизации количества не оказывают большого влияния, это служит сигналом: что-то не так. Не напрасно ли мы тратим время и деньги на эти исследования, ко-

торые вносят столь незначительный вклад в научную литературу? Низкое число цитирований не обязательно, конечно, говорит о качестве статьи. Это может быть, например, недооцененная работа. Однако если ученые публикуют бесполезные статьи только для того, чтобы получить место или грант, а не ради развития науки, то неудивительно, что столь многие из этих статей не представляют для других ученых интереса.

Триада нарезки, жульнических журналов и мошенничества с рецензиями ясно показывает, что не стоит оценивать ученых по общему числу публикаций: количество слишком уж легко подтасовать. Для решения этой проблемы предложено было оценивать ученых по числу *ссылок* на их работы. Как мы только что обсудили, этот показатель должен давать более точное представление о реальном вкладе в науку или пользе для общества. Правда, в крайнем случае ученый может написать единственную очень успешную работу с тысячами цитирований, а затем наштамповать десятки бесполезных статей, которые никто никогда читать не будет. В такой ситуации суммарное количество цитирований не будет хорошим показателем общего вклада этого ученого в науку.

В 2005 году физик Хорхе Хирш придумал способ обойти эту проблему — использовать *h*-индекс, или индекс Хирша[37]. Ученый с индексом Хирша *h* имеет *h* работ, каждая из которых была процитирована как минимум *h* раз. Это значит, что если ваш индекс Хирша равен 33, как у меня на момент написания этих строк, то у вас есть тридцать три статьи, на каждую из которых сослались не менее тридцати трех раз. Хитрость задумки в том, что увеличивать индекс Хирша становится чем дальше, тем труднее. Чтобы повысить свой до 34, мне нужно написать еще одну статью, которую процитируют по меньшей мере тридцать четыре раза, да еще и увеличить минимальное количество цитирований до тридцати четырех для других тридцати трех своих работ. Таким образом, требуется много труда — и внима-

ния со стороны других исследователей, — чтобы в итоге обзавестись индексом Хирша, равным нескольким сотням, как у некоторых выдающихся ученых. Специализированная академическая поисковая система *Google Scholar* автоматически вычисляет индекс Хирша, и многие ученые — к ним, увы, отношусь и я — немного стыдятся того, сколь часто заходят на свою страницу в *Google Scholar*, чтобы проверить, не поменялся ли индекс с ростом числа цитирований. (По моему опыту, когда ученый, даже если он высмеивает саму идею каких-либо показателей для оценки исследователей, говорит вам, что его совершенно не интересует проверка собственного индекса Хирша, он, скорее всего, либо лукавит, либо не слышал о *Google Scholar*[38].)

Как вы, наверное, догадались, индекс Хирша ученого часто в явном виде учитывается при принятии решений о приеме на работу и продвижении по службе. Таким образом, у ученых есть сильный стимул обеспечивать себе цитируемость, равно как и публиковать все больше и больше работ, которые могут быть процитированы. И опять-таки: несмотря на то, что индекс Хирша стали использовать из лучших побуждений, порождаемые им стремления могут вызывать такое поведение, какое отвечает интересам самой системы, а не целям науки.

Очевидно, что лучший способ добиться цитирования — это получить важные, революционные результаты. И некоторые исследователи, как мы уже видели, тратят неимоверно много времени на попытки убедить журналы (и весь мир) в том, что их результаты именно таковы. Раскрутка, обсуждавшаяся в предыдущей главе, помогает с цитированиями: в одном исследовании было показано, что на статьи со значимыми результатами ссылаются в 1,6 раза чаще, чем на те, что сообщают о результатах отрицательных, однако на работы, в которых авторы *открытым текстом* заключали, что результаты подтверждают их гипотезу, ссылались в 2,7 раза чаще[39]. Мораль ясна: если вы хотите, чтобы вас цитировали, пишите статью в более позитивном ключе — даже если это означает, что придется в словесной форме сгладить у своих результатов все острые, но реальные края.

Впрочем, гораздо более эффективный способ увеличивать количество цитирований — просто ссылаться на свои собственные работы. Самоцитирование — а в первые три года после публикации статьи, как было показано в одном анализе, около трети всех ссылок на нее составляют как раз цитирования самих же ее авторов — зона серая[40]. Наука развивается постепенно, и исследователи работают над конкретной темой долгие годы. Бессмысленно было бы запрещать им ссылаться на собственные предыдущие статьи, когда они делают очередной шаг в своей исследовательской программе. Однако кто-то заходит слишком далеко. Граница между приемлемым и сомнительным самоцитированием зачастую размыта, но некоторые случаи очевидны[41]. Психолог Роберт Стернберг ушел с поста редактора престижного журнала *Perspectives in Psychological Science* в 2018 году после того, как подвергся резкой критике за, помимо прочего, практику самоцитирования[42]. Проблема была вот в чем: редакторы журналов обычно пишут редакционные статьи с собственными комментариями к работам, публикуемым в данном номере. Когда Стернберг писал такие редакционные статьи, он часто нашпиговывал их ссылками на собственные работы: в семи редакционных статьях 46 % ссылок вели к его собственным работам, а в одной из них этот показатель достигал 65 %[43]. Поскольку как редактор вы отвечаете за то, что будет опубликовано в журнале, требуется определенный самоконтроль, чтобы не злоупотреблять своим положением ради повышения собственного индекса Хирша. У некоторых редакторов такого самообладания, похоже, поменьше, чем у других.

Если подобное накачивание собственного индекса Хирша кажется вам чересчур откровенным, есть вариант заставить других делать это за вас. Почти любой ученый расскажет вам о случае, когда анонимный рецензент невзначай порекомендовал сослаться на работы X, Y и Z, по случайности имеющие, как ни странно, одного общего автора — автора, который, конечно же, никак не *может* быть тем самым анонимным рецензентом. Помимо таких рассказов есть и до-

казательства: в исследовании рецензий, содержащих “предложения” по цитированию, было показано, что 29 % навязанных ссылок вели к собственным работам рецензента, а еще — что предложения процитировать статьи рецензента чаще встречались в положительных рецензиях, чем в отрицательных (то есть рецензенты чаще предлагали ссылаться на их собственные работы в рецензиях на те статьи, которые рекомендовали к публикации)[44].

Кроме того, Роберт Стернберг занимался своего рода гибридом нарезки и самоцитирования — самоплагиатом. В новых работах он повторно использовал фрагменты текста, которые ранее где-то уже публиковал. Казалось бы, как можно заниматься самоплагиатом? Разве суть плагиата заключается не в том, что идеи и формулировки крадутся у других людей? Вторичное использование текста — это, конечно, дармоедство, но оно, по крайней мере, не увеличивает количество неудачных или ложных идей в мире. Однако самоплагиат нарушает авторский договор — иногда буквальный, в случае с авторскими правами, но главное, метафорический, заключенный с читателем, — предусматривающий, что работа является оригинальной. Пережевывание одних и тех же идей зачастую позволяет ученым выглядеть более продуктивными и, подобно нарезке, создает неравные условия для сравнения резюме исследователей.

В последние годы многих ученых уличили в повторном опубликовании значительных объемов текста, иногда даже целых статей, в нескольких журналах без какого-либо указания на то, что они это делают. В одном из случаев, касающихся Стернберга, он взял статью, опубликованную им в “Журнале когнитивного образования и психологии” *(Journal of Cognitive Education and Psychology)*, объединил ее с кусками текста из главы еще раньше изданной книги, изменил “когнитивное образование” в названии на “школьную психологию” и опубликовал в “Международном журнале школьной психологии” *(School Psychology International)*[45]. Редактор этого

журнала в конце концов статью отозвал — как "дублирующую публикацию"[46]. В одной работе, посвященной анализу статей довольно небольшой выборки австралийских ученых, обнаружилось, что если самоплагиат определить как повторное использование десяти или более процентов текста из предыдущих статей в последующих работах без соответствующих указаний, то шесть из десяти авторов, попавших в выборку, в нем повинны[47].

Многие из методов, к которым, как мы видели, прибегают ученые, чтобы перехитрить систему публикаций и цитирования, могут быть использованы и журналами. Это особенно удручает, ведь журналы должны быть гарантами научных стандартов. Вот еще одно доказательство, что проблемы, с которыми сталкивается наука, комплексные, то есть наперекосяк пошла вся система.

Эквивалент индекса Хирша для журналов называется *импакт-фактором*. Изначально он задумывался как инструмент, который помогал бы университетским библиотекарям, располагающим ограниченным бюджетом, выбирать, на какие журналы оформлять подписку[48]. Однако со временем он превратился в официально признанную количественную оценку важности и престижности любого журнала. В целом импакт-фактор, рассчитываемый ежегодно, представляет собой среднее количество цитирований, которое набирают за конкретный год последние статьи из журнала[49]. На момент написания этих строк импакт-факторы для супервысокостатусных *Nature* и *Science* равны 43,070 и 41,063 соответственно; журналы, находящиеся внизу издательской иерархии, имеют импакт-факторы, представляющие собой десятичную дробь с целой частью из единственной цифры[50].

Таким образом, импакт-фактор — это среднее значение, но, поскольку различные статьи, опубликованные в одном и том же журнале, иногда имеют совершенно разную судьбу,

это может быть средним значением крайне широкого набора чисел. Распределение цитирований очень похоже на распределение доходов: нескольким статьям на вершине достается львиная доля, а большинству — относительно мало, если вообще не ноль[51]. Самые высокоцитируемые статьи и тянут средний показатель вверх, а значит, не стоит рассчитывать на то, что ваша свеженькая статья в *Nature*, несмотря на импакт-фактор этого журнала, наберет в ближайшем будущем сорок три цитирования, как нельзя ожидать, что заработок большинства людей, встреченных вами в той или иной стране, равен среднему для этой страны доходу.

Тем не менее в нынешней системе чем больше импакт-фактор журнала, тем лучше для его репутации. Большинство журналов издается коммерческими компаниями, такими как *Elsevier* и *Springer*, поэтому неудивительно, что редакторы могут оказаться под сильным давлением со стороны издателей, заставляющих повышать импакт-фактор — даже в нарушение правил научной этики. Некоторые редакторы используют трюк Стернберга, набивая свои редакционные статьи ссылками, которые “по случайному совпадению” ведут только к работам, опубликованным в этом же самом журнале, причем только к тем, что вышли за последние два года, ведь ежегодный перерасчет импакт-фактора не учитывает публикации старше[52]. Какие-то редакторы действуют подобно рецензентам, которых мы обсуждали выше: склоняют авторов к так называемому принудительному цитированию, то есть во время рецензирования требуют, чтобы те сослались на целый перечень статей, ранее опубликованных в этом журнале, независимо от того, имеют те работы непосредственное отношение к данной или нет. В одном опросе около 20 % ученых заявили, что такое с ними происходило[53].

Бывает даже, что редакторы создают “картели цитирования”, где между несколькими разными журналами заключаются тайные соглашения о цитировании статей. Один вопиющий случай раскрыл в 2012 году специалист по издатель-

скому делу Фил Дэвис, и из его рассказа о том, как он следовал по тропе цитирований вниз по извилистой кроличьей норе, стоит привести объемную выдержку:

> В 2010 году в журнале *Medical Science Monitor* вышла обзорная статья, содержащая 490 ссылок, причем 445 из них — на статьи, опубликованные в журнале *Cell Transplantation*. Все 445 ссылок вели к работам, вышедшим в 2008 или 2009 годах… Из оставшихся 45 ссылок 44 вели к статьям из журнала *Medical Science Monitor*, опять же опубликованным в 2008 и 2009 годах.
> Трое из четырех авторов этой статьи входят в состав редакционной коллегии журнала *Cell Transplantation*. <…>
> В том же 2010 году двое из этих редакторов также опубликовали обзорную статью в *The Scientific World Journal* со ссылками на 124 работы, 96 из которых были опубликованы в журнале *Cell Transplantation* в 2008 и 2009 годах. Из оставшихся 28 ссылок 26 вели к статьям, опубликованным в *The Scientific World Journal* в 2008 и 2009 годах. Прослеживается закономерность[54].

В свете растущего числа картелей цитирования компания *Thomson Reuters*, которая рассчитывает импакт-факторы, начала исключать определенные журналы из своих рейтингов из-за их практики "аномального цитирования"[55].

Таким образом, как и с количеством публикаций и индексом Хирша, с импакт-фактором тоже можно мухлевать. И как только ученые начинают за счет самоцитирования, принудительного цитирования и других подозрительных методов искусственно раздувать подобные показатели, те перестают быть мерилами научного качества. Они начинают меньше говорить о том, какие ученые и какие журналы лучшие, и больше о том, какие из них наиболее целенаправленно работают над повышением своих показателей. Закон Гудхарта налицо: "Когда

мера становится целью, она перестает быть хорошей мерой"[56]. Как мы видим, эти меры в нашей современной научной культуре превратились, по сути, в явные цели, вызвав непредвиденные последствия — создав систему порочных стимулов, которая поощряет бессмысленные показатели и дутые рейтинги, а не воспроизводимость, строгость и подлинный научный прогресс.

Особенно обескураживает то, что люди, запутавшиеся в этих дебрях бесполезных цифр, — ученые, ведь предполагается, что именно они лучше других разбираются в статистике и наиболее критичны к злоупотреблениям ею. И все же каким-то образом оказывается, что они работают в системе, где эти пустые и вводящие в заблуждение показатели ценятся превыше всего. Поначалу кажется, что с научной точки зрения удобно иметь числа, которые помогают количественно оценить вклад ученого или журнала; в конце концов, объективная количественная оценка — это одна из уникальных сильных сторон науки. Но, как гласит закон Гудхарта, чуть только вы начинаете гнаться за самими цифрами, а не за стоящими за ними принципами — в данном случае за принципом провести исследование, которое существенно увеличит наше общее знание, — вы совершенно сбиваетесь с пути. Тот факт, что одержимость этими показателями не только является уделом отдельных ученых, борющихся за статус, а вплетена в саму ткань как университетской, так и публикационной системы, еще раз подтверждает, насколько очевидно научная система не справляется со своей главной задачей.

На страницах этой книги мы обсуждали множество факторов, ведущих к плохим исследованиям. Есть ученые, которые так жаждут почувствовать, что принесли пользу, и настолько обманываются своими собственными теориями, что прибегают к мошенничеству или *p*-хакингу ради устранения любых беспокоящих неясностей. Есть ученые, кото-

рыми движет в первую очередь страсть к деньгам, престижу, власти или славе и которым на истину, как и любому шарлатану, плевать. Есть ученые слишком занятые или задерганные, чтобы проверять, нет ли в их работе ошибок. Есть те, кто не хотят подвергать сомнению когда-то полученные знания и продолжают использовать старые ошибочные методы. Правильно ли считать систему научных публикаций некой *ур*-проблемой*, лежащей в основе всего перечисленного? Вправе ли мы сказать, что порочные стимулы, созданные приоритизацией публикаций, цитирований и грантовых денег, *непосредственно* привели к мошенничеству, предвзятости, недобросовестности и хайпу?

Нам никогда не узнать наверняка, что совершается в голове у ученого, когда он применяет один из тех сомнительных методов, которые мы рассматривали в предыдущих главах. Но мы можем попытаться вывести наилучшее объяснение. Ученые — люди, а люди реагируют на стимулы. Обсуждавшиеся нами проблемы в науке так широко распространены во всем мире и во всех научных областях, что они должны иметь объяснение на уровне научной культуры в целом; мы говорим не просто о нескольких паршивых овцах, разрушающих науку для всех остальных. Когда мы видим общие тенденции в научной практике последних десятилетий — экспоненциальный рост количества статей, сильный академический отбор по публикациям, цитированиям, индексам Хирша и грантам, одержимость импакт-факторами и новыми захватывающими результатами и возникновение таких явлений, как хищнические журналы, которые, конечно же, просто удовлетворяют спрос, — разве не странно было бы при этом *не* наблюдать подобного плохого поведения со стороны ученых? Хотя нам не стоит прекращать поиски других объяснений — возможно, например, что проблема кроется скорее в контролировании системы, а не в са-

* От немецкой приставки *ur*-, обозначающей “первоначальный”, “исходный”.

мих стимулах, — предположение, согласно которому именно стимулы, связанные с публикациями, ведут к деградации науки, довольно неплохо объясняет то состояние, в коем мы оказались. Во всяком случае, мы можем сказать, что такой мотив подходит ко всем преступлениям.

Чрезвычайно трудно провести метанаучный эксперимент над всей системой, учитывая, что он должен был бы охватить целые карьеры и тысячи университетов и журналов в разных странах и в разных областях исследований. И тем не менее мы в силах сделать больше, чем просто строить догадки, и некоторые изобретательные ученые разработали компьютерные модели, в которых имитируется публикационная система для изучения того, как ее стимулы влияют на исследования.

Какие-то из этих моделей рассматривают научную систему в терминах эволюции. Выше я сравнил процесс удлинения резюме, требуемых для получения академических должностей, с половым отбором, когда для привлечения партнеров развиваются все более экстравагантные признаки. Однако можно провести и другую эволюционную аналогию. Как мы уже видели, научная система сейчас устроена таким образом, что поощряет тех, кто использует нечестные методы. Раз порядочные исследователи — кто работает во имя науки, а не ради статуса, денег или для достижения других ненаучных целей — в этой системе неконкурентоспособны, они чаще будут покидать академический мир и искать какую-то другую работу. По крайней мере, они будут менее конкурентоспособны в борьбе за высшие должности. А значит, помимо того, что система подталкивает всех к неблагонадежной исследовательской практике, она еще и отбраковывает ученых, которые твердо намерены делать все честно, замещая их теми, кто с удовольствием отступает от правил.

Модель, построенная специалистом по когнитивной науке Полом Смалдино и экологом Ричардом Макэлритом, пожалуй, ярче всего иллюстрирует, как этот процесс может выглядеть в динамике[57]. Все напоминает игру с не-

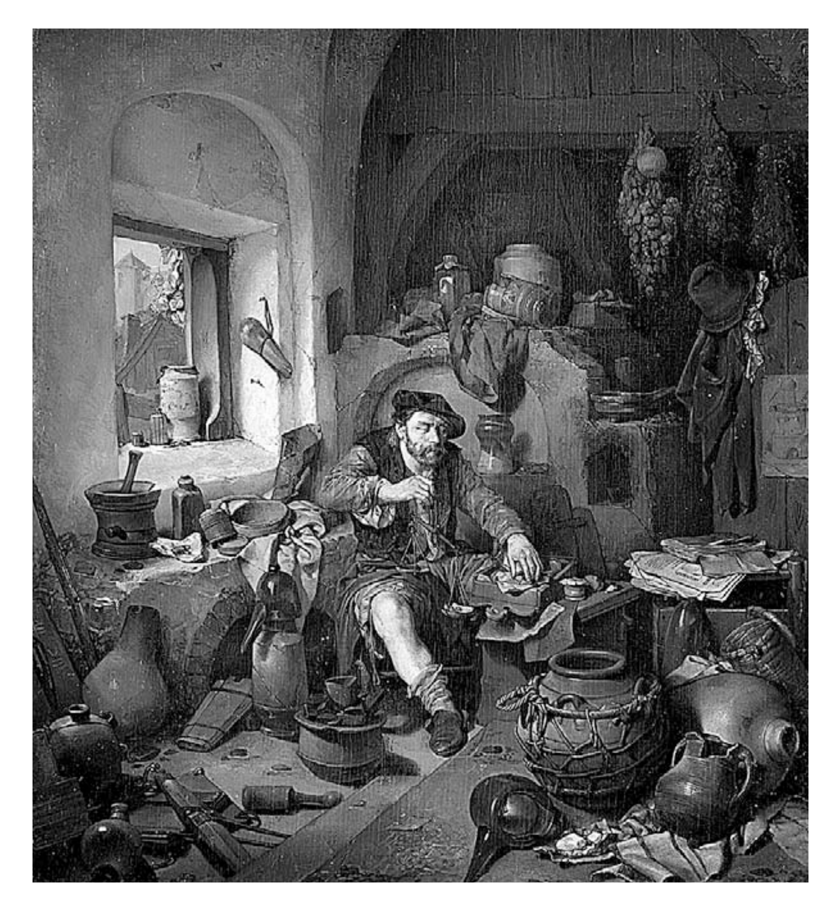
сколькими ходами. Исходно есть несколько лабораторий, и в каждой ученые проверяют новые гипотезы, с разной степенью тщательности стараясь предотвратить появление ложноположительных результатов, а затем пытаются опубликовать статью. Если лаборатория получает положительные результаты, то вознаграждается публикацией; если ее эксперимент ничего не дал, вознаграждения нет. В этой модели на каждом этапе лаборатории с бóльшим числом публикаций чаще "воспроизводятся", то есть посылают своих благополучно взращенных аспирантов создавать собственные лаборатории, тем самым распространяя в научном сообществе свои методологические приемы (и уровень их тщательности). По мере развития модели стимулы оказывают свое тлетворное воздействие: вознаграждение виртуальных лабораторий "воспроизводством", если они публикуются чаще, означает, что все больше и больше лабораторий прилагают все меньше и меньше усилий к тому, чтобы обеспечивать высокое качество своей науки. Это происходит потому, что, как ни странно, ложноположительные результаты опубликовать не сложнее истинно положительных, зато легче получить. В итоге количество опубликованных ложных открытий стремительно растет. Смалдино и Макэлрит называют это "естественным отбором плохой науки".

Другие компьютерные модели, имитирующие публикационную систему, дают аналогичные результаты. Одна из них показала следующее: учитывая, что в научной литературе решительное предпочтение отдается новизне, оптимальная стратегия для остервенело честолюбивого ученого — "провести множество небольших исследований с малой статистической мощностью, чтобы максимизировать количество публикаций, даже если это означает, что около половины дадут ложноположительные результаты"[58]. Другая модель продемонстрировала, как тяга научных журналов к положительным результатам привела к "извращенному поощрению ложноположительных и мошеннических результатов в ущерб добро-

совестной науке"[59]. Разумеется, компьютерные модели — это не реальность, где различных параметров значительно больше. Однако такие упрощенные симуляции математически подкрепляют сделанные мною выше умозаключения, показывая, как недостатки системы стимулирования могут со временем снизить качество науки.

В Музее Гетти в Лос-Анджелесе есть картина золотого века голландской живописи — "Алхимик" кисти Корнелиса Бега[60]. В хаосе своей лаборатории сидит алхимик, вокруг него надтреснутые горшки, вазы с обитыми краями и расколотые реторты — плоды неудавшихся попыток превратить неблагородные металлы в золото. Вопреки распространенному мнению, алхимия не была настолько уж никчемной, и грань между некоторыми видами алхимической деятельности и ранними проявлениями того, что мы сейчас величаем химией, по меньшей мере туманна[61]. Но Бега живописал тщетность одержимости золотом. И это хорошая аналогия для современной системы стимулов в науке. Погоня за академическими сокровищами, такими как публикации и цитирования, оставила нас с обломками никуда не годных научных исследований[62].

Порочные стимулы действуют как злой джинн, давая вам именно то, что вы просили, но не обязательно то, чего вы хотели. Поощряйте наращивание числа публикаций, и вы его добьетесь — но будьте готовы к тому, что у ученых останется меньше времени на проверку ошибок, а нарезка статей станет нормой. Поощряйте публикации в журналах с высоким импакт-фактором, и вы их получите — но будьте готовы к тому, что ученые в своих попытках пробиться туда станут прибегать к *p*-хакингу, публикационному смещению и даже мошенничеству. Поощряйте конкуренцию за гранты, и вы ее добьетесь — но будьте готовы к тому, что ученые начнут без всякой меры раздувать свои результаты и подавать



Корнелис Бега, “Алхимик”, 1663 г. Музей Гетти

их под нужным углом в попытке привлечь внимание спонсоров. На первый взгляд может показаться, будто наша нынешняя система финансирования науки и публикации результатов поощряет продуктивность и новаторство, однако вместо этого она часто вознаграждает тех, кто следует только букве, а не духу закона науки[63].

Осознание проблемы со стимулами не означает, что ученым нужно спускать с рук недобросовестные действия. Все

мы чувствуем силу стимулов, однако должны делать все возможное, чтобы противостоять им, — во имя науки[64]. Правда, было бы лучше, если бы нам не приходилось, пытаясь делать открытия о мире, сопротивляться сокрушительной тяжести системы "публикуйся или погибни". Было бы лучше, если бы мы сумели достичь золотой середины, когда поощряется усердная работа и изобретательность ученых, но еще и осторожность со строгостью; когда ученые стремятся сделать все *правильно*, а не просто опубликоваться[65]. Так как же нам этого достичь? Как нам выправить наши стимулы и тем самым повысить надежность науки? Мы обсудим это в следующей, заключительной главе[66].

Глава 8
Исцеление науки

> Процесс научного открытия…
> за следующие двадцать лет изменится
> сильнее, чем за последние триста.
>
> Майкл Нильсен [1]

Большинство обсуждавшихся в этой книге проблем с наукой отражено в одной метанаучной работе 2018 года. Исследовательница психиатрии Имкье Анна де Врис и ее коллеги изучили все этапы, которые обычно отделяют клинические испытания нового лекарства от его финального представления миру[2]. В свою выборку они включили сто пять различных испытаний антидепрессантов, которые были одобрены Управлением по контролю качества пищевых продуктов и лекарственных средств США. Соотношение положительных и отрицательных результатов оказалось почти пятьдесят на пятьдесят: в пятидесяти трех исследованиях было показано, что рассматриваемый антидепрессант работает лучше, чем контроль или плацебо, а в пятидесяти двух результаты были либо отрицательными, либо неоднозначными[3]. Пока все предсказуемо: некоторые исследования дали значимые результаты, а некоторые — нет. Проблема в том, что произошло дальше.

Просматривая все исследования, де Врис и ее команда наблюдали процесс “отмывания литературы”, превративший беспорядочный, пестрый набор испытаний в куда более четкую историю научного открытия, — процесс, благодаря которому рассматриваемые препараты стали выглядеть значительно более эффективными, чем они на самом деле были. Первым шагом стало публикационное смещение. В конеч-

ном итоге были опубликованы 98 % исследований с положительными результатами (пятьдесят два из пятидесяти трех), тогда как с отрицательными — только 48 % (двадцать пять из пятидесяти двух). В реальности соотношение положительных и отрицательных результатов было сбалансированным, близким к единице, а после публикации положительные результаты вдвое превосходили отрицательные. Как мы знаем, в научных журналах отрицательные результаты не приветствуются, однако для ясного видения научной летописи нам необходимо о них знать.

Вторым шагом стал *p*-хакинг, а именно переключение на другие исходы, при котором ученые смещают фокус своего исследования, когда обнаруживается, что их основной результат не является статистически значимым. Как только изменились изучаемые исходы, еще десять исследований с отрицательными результатами превратились в положительные (помните, в четвертой главе мы говорили, как переключение на другие исходы увеличивает вероятность, что наш результат — всего лишь ложноположительный). Пока еще оставалось пятнадцать исследований с явно отрицательными результатами — и в дело вступила подача под нужным углом. В десяти из этих пятнадцати исследований в аннотации или основном тексте статьи для придания результатам более позитивного характера в той или иной форме использовалась подача под нужным углом.

После этих хитростей осталось всего пять однозначно отрицательных исследований — то есть вдесятеро меньше, чем было исходно. И гнилой вишенкой на торте послужило то, что в итоге положительные испытания цитировались в последующих работах в три раза чаще, чем отрицательные. Весь этот удручающий процесс пошагово показан на рисунке 4.

Это характерно не только для исследований антидепрессантов: авторы отследили аналогичную цепочку событий в испытаниях новых методов психотерапии[4]. На самом деле то же самое происходит — в большей или меньшей сте-

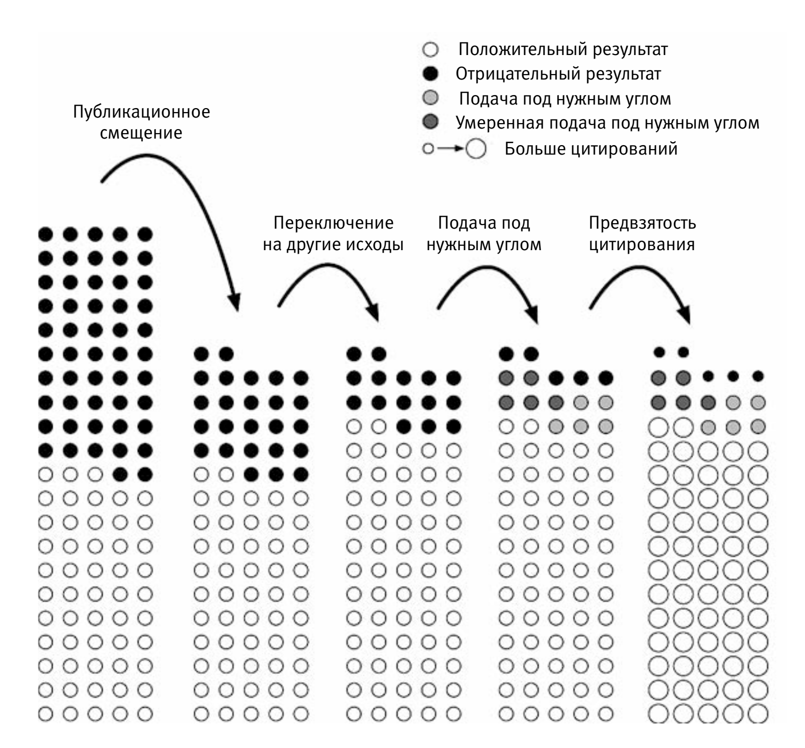


Рис. 4. Хитрая схема, скрывающая отрицательные результаты. Каждая точка представляет собой испытание антидепрессантов; черные точки — испытания с отрицательными результатами. Мы можем наблюдать, как по мере движения слева направо исследования с отрицательными результатами исчезают. Отрицательные исследования публикуются значительно реже, чем положительные; превращаются в положительные за счет переключения на другие исходы; подаются под таким углом, чтобы выглядеть положительными, слегка либо значительно; цитируются гораздо реже, чем положительные исследования, чьи точки в последней колонке раздуты. В итоге мы видим почти исключительно положительные результаты. Адаптировано из De Vries Y. A. et al., 2018.

пени — почти со всеми научными исследованиями. Любой будущий специалист по метаанализу, который захочет объять полную картину конкретной области исследований, получит чудовищно искаженную перспективу — и даже не узнает

об этом, если, конечно, не изучит подробно регистрационные документы клинических испытаний (которые, кстати, не являются обязательными ни в каких других областях, кроме медицины). В работе де Врис не изучалось возможное мошенничество, или ошибки в дизайне исследования и при анализе результатов, или раскрутка при представлении методов лечения в средствах массовой информации и их продвижении, но мы без колебаний можем предположить, что и это все постоянно происходит, еще больше размывая реальные исследования.

Очевидно, что вскрытое де Врис и ее коллегами далеко от описанного в начале этой книги идеала науки, предполагающего, что социальные процессы публикации и рецензирования обеспечивают распространение по миру фактов и отсеивание любых случайных ошибок. Так как же науку спасти? Как нам перестроить культуру и стимулы для процессов публикации научных результатов, финансирования науки и определения приоритетов, чтобы они поощряли истину и надежность вместо ярких историй? Как нам оградить науку от ученых (воспользуемся названием документальной программы на радио BBC)?

В этой заключительной главе мы в общих чертах обсудим ответы на подобные вопросы. Во-первых, мы рассмотрим широкий спектр изменений, которые можно внести, чтобы предотвратить четыре основные проблемы: мошенничество, предвзятость, недобросовестность и хайп — или по крайней мере смягчить их последствия. Затем мы обсудим способы, какими можно преобразовать не только повседневную работу ученых, но и саму научную культуру. Некоторые изменения уже происходят, другие — дерзновенные проекты, способные произвести революцию в том, как наука делается.

В третьей главе мы видели, что университеты часто экранируют научных мошенников от последствий их действий. В знаменитых случаях, которые мы рассмотрели, плотину се-

кретности в конце концов прорвало и имена мошенников всплыли: Дидерик Стапел, Паоло Маккиарини, Хван У Сок, Ян Хендрик Шён и легион других. Но во многих случаях менее злостного нарушения норм научной работы имена так никогда и не предаются огласке[5]. Неудивительно, что вокруг безнаказанно промышляет столько научных мошенников, если мало шансов, что их заметят, *а кроме того*, мало шансов, что, если их все-таки поймают, многие о том узнают. Таким образом, первое, что стоит менять: нужно хотя бы просто называть и стыдить тех, кого уличили в недобросовестной научной практике[6].

Конечно, университеты не сильно заинтересованы в том, чтобы это делать, поэтому следующая идея — отдать бразды правления в этом вопросе независимым организациям. Тут уже есть некоторые подвижки. Например, увидев, как Каролинский институт осрамился из-за дела Маккиарини о трансплантации трахей, шведское правительство в 2019 году приняло закон, запрещающий университетам самостоятельно расследовать случаи нарушения норм научной работы и предписывающий передавать эту ответственность независимому правительственному агентству[7].

Все это прекрасно, когда мошенничество уже обнаружено, но лучше было бы вообще предотвращать публикацию жульнических статей в журналах. И вот тут нам могут помочь технологии. Исследователи разрабатывают все более эффективные алгоритмы, способные выявлять в научных статьях поддельные данные и отыскивать нарушения вроде дублирования изображений[8]. Интересно будет сравнить технологические умения с навыками опытного человека, распознающего дублирования изображений, вроде Элизабет Бик из третьей главы, но теоретически алгоритмы должны сделать задачу выявления неправомерных манипуляций с данными гораздо менее трудоемкой. Журналы могли бы требовать, чтобы каждая представленная на рассмотрение работа оценивалась подобными алгоритмами (а также другими,

о которых мы говорили ранее, например GRIM и *statcheck*), дабы все подозрительное было найдено еще до начала рецензирования. Еще журналы могли бы использовать специальные программы, выискивающие в тексте любой потенциальный плагиат или самоплагиат[9].

Подобные алгоритмы также помогают бороться с небрежностью[10]. Существенная часть огрехов, выявленных алгоритмом *statcheck*, — похоже, банальные ошибки, возникшие, когда исследователи копировали числа из программ статистического анализа и вставляли в текстовый редактор, который использовали для написания статьи. Прогоняя готовый текст статьи через *statcheck*, можно было бы отлавливать такие ошибки до того, как они попадут в литературу. Однако технологии могли бы помочь предотвращать даже появление подобных ошибок. В последние годы было разработано программное обеспечение, которое объединяет статистический анализ и обработку текста в одной программе, автоматически заполняя все необходимые для статьи таблицы и создавая необходимые иллюстрации[11]. Данные могут ускользнуть от внимания отвлекающегося и ошибающегося ученого, но, поскольку вся цепочка действий от данных до статьи находится у всех на виду, намеренное манипулирование цифрами или анализом тоже становится сложнее[12].

Но не стоит этими новыми технологиями обольщаться — любое программное обеспечение подвержено своим собственным ошибкам. Приведем один особенно постыдный пример: выяснилось, что приблизительно в 20 % статей по генетике, в которых использовалась электронная таблица *Microsoft Excel* для перечисления изученных генов, из-за автозамены такие названия генов, как *SEPT2* и *MARCH1*, преобразовывались в даты[13]. Результаты работы автоматического программного обеспечения должны тщательно проверяться людьми, как минимум до тех пор, пока мы не будем уверены, что расправились со всеми “глюками”. В теории, однако, кажется, что многие рутинные научные задачи могли бы акку-

ратнее выполняться нечеловеческим разумом: анализ больших массивов данных для выявления закономерностей; перелопачивание чисел в научной литературе для выработки консенсуса; даже интерпретация изображений блотов, клеток и сканов мозга. Учитывая огромное количество ошибок, встречающихся в научных работах, и то, как легко можно было бы избежать многих из них с помощью более автоматизированного процесса написания статей, в конечном итоге может оказаться неэтичным полагаться при выполнении этих задач только на людей.

Одна из проблем, на которую мы постоянно наталкивались, — извечное пристрастие ученых к новизне. Хотя новые, захватывающие результаты движут научный прогресс, мы видели, как одержимость "революционными" открытиями привела к тому, что целые области исследований основываются на шатких, невоспроизводимых доказательствах. Перефразируя слова биолога Оттолин Лейзер, можно сказать, что смысл революционности — в непременном строительстве нового; если вы только и делаете, что разрушаете все старое, в итоге вы останетесь ни с чем[14]. Как нам сделать надежные результаты приоритетнее новых? Как побороть публикационное смещение, гарантируя публикацию всех результатов — неважно, новаторские они или отрицательные?

Одно из предложений таково: нужно создать журналы, специализирующиеся на публикации отрицательных результатов, обеспечив тем самым более привлекательную альтернативу "картотечному ящику". Так, например, в 2002 году именно с этой целью был создан журнал "Отрицательные результаты в биомедицине" *(Journal of Negative Results in Biomedicine)*. Задумка хорошая, но, наверное, неудивительно, что едва ли кто-то жаждал, чтобы его исследование вышло в низкостатусном журнале отрицательных результатов, который "определяли как издание, публикующее статьи, кото-

рые ни один другой журнал не примет"[15]. Журнал закрылся в 2017 году — необычная судьба для научного издания в мире, стонущем под грузом новых статей[16].

Если идея со специальными журналами для отрицательных результатов не сработает, то как насчет журналов, открыто принимающих любые результаты при условии, что исследование, в котором они получены, признано методологически обоснованным? Еще такие издания могли бы стать пристанищем для исследований, воспроизводящих какие-то предыдущие работы, поскольку подобные исследования страдают от тех же предрассудков, что и отрицательные результаты. В последние годы появился целый ряд таких журналов (часто называемых мегажурналами, поскольку отсутствие требования подавать именно положительные или "захватывающие" результаты означает, что там публикуется *много* материалов), включая PLOS ONE, где я и мои коллеги в конечном итоге опубликовали наш отчет о неудавшемся повторении исследования Дэрила Бема по экстрасенсорике[17]. Это явный прогресс, но такие журналы все еще рискуют восприниматься учеными, заботящимися о своем статусе, как низкоуровневые. В идеале мы хотим видеть и в знаменитых журналах с высоким импакт-фактором верную долю отрицательных результатов и больше работ, посвященных попыткам воспроизвести более ранние исследования.

И тут есть хорошие новости. Хотя и не стремясь принимать к рассмотрению отрицательные результаты, многие известные журналы сейчас смягчают свое прежнее отношение к публикации исследований-повторений. Возьмем, к примеру, *Journal of Personality and Social Psychology*, который опубликовал работу Бема и отклонил нашу из-за своей общей политики "никаких повторений". Теперь, после кризиса, на их сайте есть целый раздел, посвященный воспроизведению исследований, где отмечается: редакционная коллегия "признает важность повторений для создания совокупной базы знаний в нашей области. Поэтому мы призываем

подавать на рассмотрение работы, в которых предпринята попытка воспроизвести важные результаты, особенно из тех исследований, что были ранее опубликованы в *Journal of Personality and Social Psychology*"[18]. Это хороший пример нового правила для научных журналов, предложенного психологом Санджаем Шриваставой и основанного на надписи "Сломали — покупаете" из некоторых магазинов керамики: если вы публикуете статью, то по крайней мере частично несете ответственность за публикацию последующей работы, проверяющей, воспроизводятся ли ее результаты[19].

Все больше и больше редакторов из самых разных областей следуют этому правилу; свыше тысячи журналов приняли недавно набор руководящих принципов, прямо заявляющих, помимо прочего, что исследования-повторения приветствуются[20]. Некоторые финансирующие структуры, такие как Нидерландская научно-исследовательская организация, вливают деньги в повторение исследований[21]. Шаги позитивные, но практика покажет, действительно ли журналы начнут регулярно публиковать больше исследований-повторений. Специалисты по метанауке будут за этим наблюдать.

Если ученым проще будет публиковать повторно проведенные исследования и отрицательные результаты, публикационное смещение может уменьшиться. Но как насчет других форм предвзятости, имеющих отношение к p-хакингу, с которыми мы столкнулись? Много десятков статей и даже целые книги посвящены подводным камням p-значений: они трудны для понимания, не говорят нам того, что мы действительно хотим узнать, и ими легко злоупотреблять[22]. В этой критике есть доля правды. В целом необходимо меньше внимания уделять *статистической* значимости — p-значениям ниже произвольного порога в 0,05 — и сосредоточиться на *практической*. В исследовании с достаточно большим размером выборки (и достаточно высокой стати-

стической мощностью) даже очень маленькие эффекты — например, от таблеток, уменьшающих головную боль на 1 % от одного балла по нашей шкале боли от 1 до 5, — могут показать себя статистически значимыми, часто с *p*-значениями намного ниже 0,05, хотя по абсолютной величине эти эффекты практически бесполезны. Экономисты Стивен Зилиак и Дейрдре Макклоски писали, что это "безразмерный взгляд на статистическую значимость", когда у ученых развивается полная концентрация на *p*-значениях в ущерб рассмотрению "силы" эффекта[23].

Для решения проблемы чаще всего предлагается попросту отказаться от идеи статистической значимости. В 2019 году более восьмисот пятидесяти исследователей подписали открытое письмо в журнале *Nature* как раз об этом. "Статистической значимости пора уйти", — говорилось там[24]. Вместо того чтобы подчеркивать значимость, утверждали они, ученым следует трезвее оценивать неопределенность своих открытий, сообщая о пределах погрешности для каждого числа, и вообще быть скромнее относительно того, что возможно вывести из зачастую расплывчатых статистических результатов[25]. Тут есть здравое зерно, хотя следует иметь в виду, что для оценки погрешности чаще всего рассчитывается так называемый доверительный интервал, а он просто обеспечивает другой взгляд на данные по сравнению с *p*-значением, но не дает много новой статистической информации[26]. Однако в ванне статистической значимости есть еще и ребенок. Предоставляя ученым объективную меру, пусть и произвольную, она связывает им руки. Отказ от *p*-значений не обязательно улучшит положение; на самом деле из-за введения иного источника субъективности ситуация может серьезно ухудшиться[27]. Джон Иоаннидис полушутливо отметил, что если мы избавимся от всех подобных объективных мер, то создадим ситуацию, когда "вся наука станет похожа на эпидемиологию питания", — а это действительно пугающая перспектива[28].

Такая же критика часто звучит в отношении другой важной альтернативы *p*-значениям — байесовской статистики. Основанный на теореме о вероятности, которую сформулировал статистик XVIII века Томас Байес, этот метод позволяет исследователям учитывать силу предыдущих доказательств — называемых “априорными” — при оценке значимости новых результатов. Например, если кто-то скажет вам, что прогноз погоды предсказывает дождливый день в Лондоне осенью, этому человеку не придется долго вас убеждать. С другой стороны, если прогноз предсказывает снежную бурю в пустыне Сахара в июле, вы, вероятно, воспримете сие утверждение весьма скептически, учитывая весь предыдущий опыт, говорящий нам о знойном сахарском лете. Приверженец байесовского подхода может включить все эти предварительные свидетельства в свой расчет — и в случае с Сахарой новый прогноз должен быть чрезвычайно убедительным, чтобы опровергнуть все предыдущие метеорологические данные[29]. С *p*-значениями такого не провернуть, поскольку они почти всегда рассчитываются независимо от любых предыдущих доказательств. Однако байесовская “априорность” по своей сути субъективна: все согласятся с тем, что в Сахаре жарко и сухо, но насколько сильно *до начала исследования* мы должны верить, что определенное лекарство уменьшит симптомы депрессии или что конкретная государственная политика будет способствовать экономическому росту, — вопрос весьма спорный.

Помимо учета априорных знаний байесовская статистика имеет и другие отличия от *p*-значений[30]. Например, на нее меньше влияет размер выборки: статистическая мощность тут роли не играет, поскольку байесовский подход направлен не на определение влияния конкретного набора условий, а просто на взвешивание доказательств за и против гипотезы. Еще байесовский подход, пожалуй, ближе к обычным рассуждениям людей о статистике. Приверженцы байесовского метода спрашивают: “Какова вероятность, что моя

гипотеза верна, учитывая эти наблюдения?" — и это более интуитивный подход, чем тот, что основан на *p*-значениях, когда вопрос звучит так: "Какова вероятность, что я получил бы эти наблюдения при условии, что моя гипотеза *не* верна?"[31]

Любой статистический подход имеет свои плюсы и минусы[32]. Хотя некоторые критики в этих дебатах заявляют, будто *p*-значения — корень всех зол, числовой Крысолов, который сбивает с толку вообще-то здравомыслящих ученых, крайне маловероятно, что мошенничество, предвзятость, недобросовестность и хайп, о которых мы говорили в этой книге, вдруг испарятся, если мы просто избавимся от одного статистического инструмента и возьмем на вооружение другой. Статистика сама по себе не способна решить главную проблему: исправить изъяны человеческой природы и, как следствие, научной системы. Независимо от того, какой статистический подход будет доминирующим, некоторые ученые найдут способы обойти систему, чтобы придать своим результатам более впечатляющий вид. Как мы увидим дальше, решения этих проблем должны лежать в области мотивации и культуры.

А пока, вместо того чтобы призывать исследователей полностью отказаться от статистического метода, особенно столь глубоко укоренившегося, как проверка значимости, полезно, наверное, получше объяснить ученым, что он способен показать, а что не способен, и начать использовать его немного иначе — так, чтобы избегать ошибок. Например, недавно было предложено изменить стандартный критерий для значимости с $p < 0{,}05$ на $p < 0{,}005$, то есть сильно поднять планку, которую должны преодолеть результаты, чтобы считаться интересными[33]. Учитывая нарывы, вскрытые кризисом воспроизводимости, мы должны, казалось бы, значительно осторожнее относиться к тому, что принимаем в качестве доказательств наших гипотез. Но поднятие планки опасно: если только мы не увеличим одновременно размер всех своих выборок, наши

тесты будут обладать гораздо меньшей статистической мощностью. Правда, сторонники значения 0,005 доказывают, что проблема ложноположительных результатов, которую их метод, вероятно, уменьшит, требует решения сильнее, чем проблема результатов ложноотрицательных.

Вот еще один способ борьбы со статистической предвзятостью и *p*-хакингом: полностью вывести анализ данных из-под контроля исследователей. В таком сценарии после сбора данных ученые должны будут передавать их для анализа независимым статистикам или другим специалистам, которые, как предполагается, практически свободны от конкретных предубеждений и стремлений тех, кто разработал и провел эксперимент[34]. Управлять подобной системой было бы непросто, и легко представить, как она ведет к конфликтам, когда ученые не согласны с анализом или интерпретацией, которые назначенный статистик навязал их ценным данным[35]. Но, как и в случае с некоторыми радикальными идеями для реформ, которые мы обсудим дальше, пожалуй, в небольших масштабах это все же стоило бы попробовать сделать.

В четвертой главе мы видели, что огромное количество способов, какими можно проанализировать набор данных, тоже доставляет ученым проблемы: откуда им знать, что выбранный анализ не даст как раз случайные результаты? Альтернатива беспокойству, верно ли вы выбрали конкретный анализ, — принять проблему “сада расходящихся тропок” и прогнать *все* анализы, какие только можно провести с вашим набором данных. Вы могли бы включать *и* исключать определенных участников, объединять *и* разделять определенные переменные, вводить *и* не вводить поправку на определенные спутывающие факторы — и основывать свои выводы на том, что результаты говорят вам в целом. Этой идее дали множество названий, например “анализ кривой спецификаций”, “анализ вибраций эффектов” и, мое любимое, “анализ мультивселенной”[36]. Если мы представим себе бесконечное

число параллельных вселенных, в каждой из которых вы провели анализ немного по-другому, в какой доле из них вы обнаружили бы одинаковый эффект? А в какой доле — совершенно противоположный? Все ли эти анализы в целом сходились бы к одному и тому же общему результату?

Оксфордские психологи Эми Орбен и Эндрю Пшибыльски, например, использовали анализ мультивселенной для решения животрепещущего вопроса о влиянии экранного времени на психическое здоровье молодых людей[37]. Результаты исследований в этой области постоянно раздуваются в прессе, и во многих газетных статьях и популярных книгах утверждается, что современным подросткам вредит то количество времени, которое они проводят онлайн[38]. Социальные сети видятся особой проблемой, поскольку якобы уменьшают число личных контактов подростков с другими людьми, подвергают их воздействию кибербуллинга и жесткой порнографии, а также снижают устойчивость внимания[39]. Были даже предложены новые психологические диагнозы: "видеоигровое расстройство", "зависимость от онлайн-порнографии", "зависимость от айфона", список можно продолжить[40]. Бóльшая часть доказательств, которые вызывают панику, связанную с технологиями, получена в крупных наблюдательных исследованиях, изучающих корреляции между временем, проведенным подростками за экраном, и проблемами с их психическим здоровьем. Учитывая большой потенциал для *p*-хакинга в подобных исследованиях (вспомните, как легко было к нему прибегать в исследованиях по питанию с большими массивами данных, когда практически все продукты могли оказаться так или иначе связанными с раком), они идеальные кандидаты для метода "анализ мультивселенной".

Орбен и Пшибыльски выбрали три больших набора данных наблюдений и рассмотрели все возможные способы, какими их можно проанализировать для проверки утверждений о вреде экранного времени. Например, можно считать "экранным временем" только просмотр телевидения или же

учитывать и видеоигры; можно спрашивать о самооценке и суицидальных мыслях, а не только о психофизическом состоянии или выбрать два вопроса из этих трех либо все три; можно ввести поправки на такие факторы, как гендер или школьные оценки, либо на любое количество других потенциально важных переменных; можно запросить оценки у родителей или самоотчеты либо и то и другое; можно использовать средние значения или же суммарные баллы из опросников; и так далее и тому подобное. Общее число "оправданных" комбинаций — то есть таких, при которых возможно привести правдоподобно звучащий научный аргумент, подтверждающий, что этот способ анализа данных и был правильным, — исчислялось сотнями для первого набора данных, десятками тысяч для второго и сотнями миллионов для третьего (в этом последнем случае, поскольку выполнение такого количества анализов перегрузило бы почти любой компьютер, авторы сократили количество комбинаций до "всего лишь" двадцати тысяч).

Проверив все эти комбинации, Орбен и Пшибыльски обнаружили: несколько анализов свидетельствуют о довольно существенном негативном влиянии экранного времени, некоторые вообще не показывают никакого эффекта, а какие-то демонстрируют, что экранное время, как ни странно, приносит пользу. Исследователи взяли среднее значение. Оно оказалось отрицательным, но корреляция между экранным временем и психическим здоровьем была крайне слабой: время, проведенное за экраном, отвечало примерно за 0,4 % изменений в психофизическом состоянии. Чтобы легче было оценить этот результат, скажу, что психофизическое состояние коррелирует примерно так же с регулярным употреблением в пищу картофеля и сильнее — с ношением очков. Вот вам и все страшилки. Анализ мультивселенной намекает, что, если мы хотим найти объяснение проблемам подростков с психическим здоровьем, не получится спихнуть все на экранное время как на козла отпущения[41]. Более общий вывод очеви-

ден: вместо того чтобы проводить только один-единственный анализ, который может соответствовать личным предрассудкам исследователя, мы должны смотреть на статистику гораздо шире, учитывая все альтернативные сценарии и задаваясь вопросом, что могло бы произойти, реши мы провести анализ немного иначе.

Недостаток метода “анализ мультивселенной” состоит в том, что обычно он требует работы суперкомпьютера, для большинства исследователей недоступного. И хотя анализ такого рода — отличный способ внести больше ясности в горячо обсуждаемые вопросы, он не избавляет ученых от постоянного давления и соблазна отбирать наиболее впечатляющие результаты и представлять их в качестве первоначальной гипотезы. Чтобы решить эту проблему, можно использовать другой инструмент для устранения неполадок в науке — *предварительную регистрацию*.

Предварительная регистрация обязательна для финансируемых правительством США клинических испытаний с 2000 года, а также это непременное условие для публикации в большинстве медицинских журналов с 2005-го[42]. Регистрация исследования подразумевает размещение в интернете общедоступного документа с отметками времени, где еще до сбора каких-либо данных должно быть подробно описано, что ученые планируют делать. Общедоступное хранилище экспериментальных планов служит ориентиром, по которому можно отследить, какая доля исследований действительно доходит до публикации. И оно позволяет нам увидеть, какие гипотезы ученые намеревались проверять, так что мы можем выяснить, не были ли они в середине исследования изменены.

В дополнение к предварительной регистрации задуманного исследования ученые еще могут предварительно регистрировать подробный план, по которому собираются анали-

зировать данные. Мы видели, что именно *незапланированный* характер статистического анализа — скрытая гибкость — может по расходящимся тропкам увести ученых к результатам, которые статистически значимы (и пригодны для публикации), но на деле не соответствуют реальности. Идея предварительной регистрации вашего статистического анализа — это научный пакт Одиссея: размещая план своего анализа где-нибудь в открытом доступе, вы привязываете себя к мачте и не даете себе поддаться зову сирен — *p*-хакинга.

Кто-то справедливо возразит, что если ученые совсем не будут оставлять себе пространства для маневра, то не останется и шансов совершить случайное открытие (пенициллин и виагра — вот два из наиболее известных случайных открытий, часто выдвигаемые в качестве аргумента)[43]. Однако предварительная регистрация — это совсем другое. В предварительно зарегистрированном исследовании какой-то спонтанный анализ для изучения интересных закономерностей в данных тоже разрешен, просто он не может быть подан под таким углом, словно был спланирован заранее. Этот так называемый *исследовательский* анализ может привести ко многим важным идеям и открытиям: например, вы неожиданно обнаружите, что новое лекарство лучше действует на пожилых, чем на молодых участников, а затем выстроите новую линию исследований, чтобы разобраться в причинах. Но, как мы уже неоднократно видели в предыдущих главах, числа всегда зашумлены, так что вы гарантированно найдете *что-то* любопытное, если вдоль и поперек проанализируете свои данные достаточным количеством способов. Поскольку вы дали себе больше попыток найти нечто статистически значимое, положительные результаты исследовательского анализа с гораздо большей вероятностью окажутся случайностью, которая на новой выборке не воспроизведется. И все же, что несколько возмутительно, в науке результаты *исследовательского* анализа по большей части преподносятся так, будто они *подтверждающие*, словно это результаты тестов, запланированных до начала

исследования. Предварительная регистрация позволяет вам четко сказать читателям, использовали ли вы данные в исследовательских целях, чтобы сформулировать гипотезу ("О, интересно, переменная X, похоже, связана с переменной Y! Надо бы проверить, воспроизведется ли это в новом наборе данных"), или в подтверждающих, чтобы гипотезу проверить ("Я предсказал, что переменная X будет связана с переменной Y в этом наборе данных, и это действительно так!")[44].

Исследование крупномасштабных испытаний, посвященных средствам профилактики и лечения сердечно-сосудистых заболеваний, результаты которого показаны на рисунке 5, ярко иллюстрирует влияние предварительной регистрации[45]. До того, как регистрация стала обязательной, в журналах сообщалось о впечатляющем количестве положительных результатов, изображенных белыми точками в нижней половине графика (то есть указывающих на сниженный риск изучавшихся исходов, связанных с сердечно-сосудистыми заболеваниями), а также о случайных отрицательных. Но посмотрите, что произошло после 2000 года, когда ввели обязательную предварительную регистрацию. Внезапно положительных результатов стало только два, а остальные исследования дали отрицательные результаты, группирующиеся вокруг нулевого эффекта (он соответствует единице по вертикальной оси "относительного риска"). Успешность испытаний до введения обязательной предварительной регистрации составляла 57 %, а после — рухнула до 8 %. Повторюсь: до введения регистрации испытания *казались* чрезвычайно успешными, свидетельствующими о множестве вроде бы эффективных средств от сердечно-сосудистых заболеваний, после же правда вышла наружу: лекарства и пищевые добавки, проверяемые в этих исследованиях, и близко не такие полезные, как нас уверяли.

Не стоит совершать ошибку отождествления корреляции и причинности, считая, что новые требования регистрировать исследования обязательно *вызвали* сокращение количества положительных результатов, — возможно, в том же году

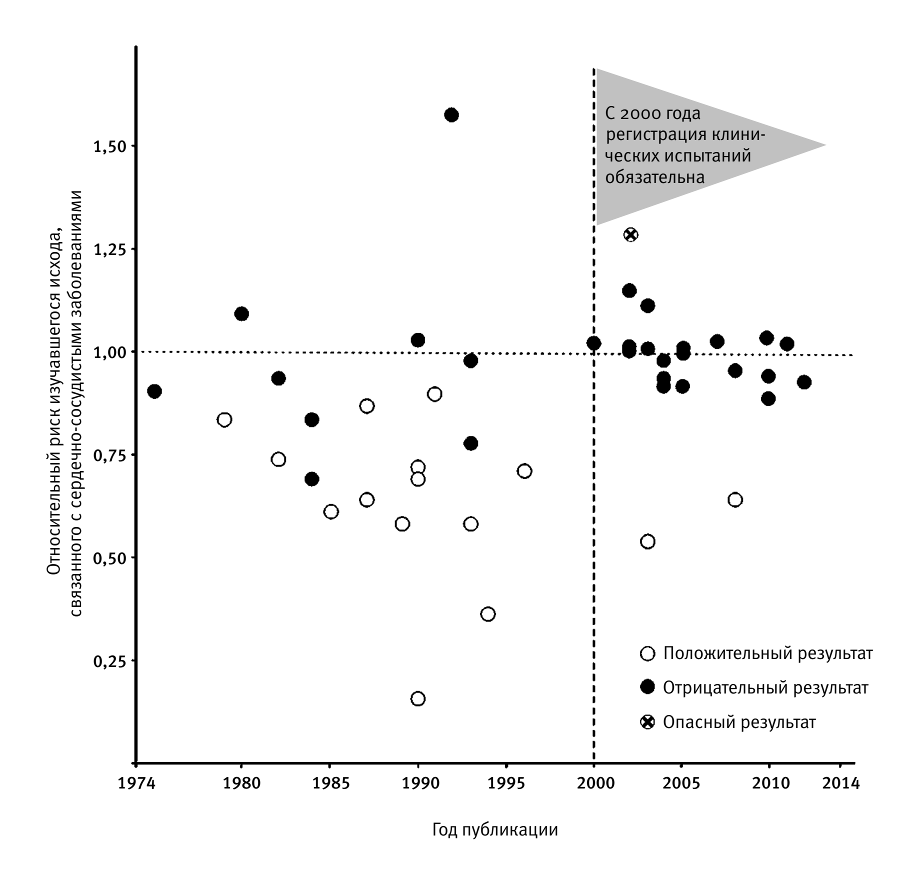


Рис. 5. Исследования, посвященные лечению и профилактике сердечно-сосудистых заболеваний и опубликованные до либо после 2000 года, когда регистрация клинических испытаний (или по меньшей мере их основных исходов) стала обязательной (вертикальная пунктирная линия). В исследованиях, расположенных ниже горизонтальной пунктирной линии, получилось, что изучаемые средства профилактики или лечения сердечно-сосудистых заболеваний снизили риски (чем ниже точка, тем меньше риск). Белыми точками обозначаются исследования, в которых обнаружился статистически значимый эффект, говорящий в пользу профилактики или лечения; черными — в которых результаты оказались отрицательными. Обратите внимание, что еще есть одно исследование со статистически значимыми результатами, проведенное после 2000 года, в котором обнаружилось, что изучаемое средство лечения вообще опасно (точка с крестиком). Адаптировано из Kaplan R. M., Irvin V. L., 2015.

произошли и какие-нибудь иные изменения, например сдвиг фокуса на другие виды лечения. Но мы вправе предположить, что публичная предварительная регистрация своих планов могла побудить ученых, проводящих клинические испытания, быть более честными и откровенными в отношении того, что они обнаружили. Если результаты свидетельствуют все-таки о причинно-следственной связи, то они служат сильным аргументом в пользу предварительной регистрации всех исследований вообще, а также суровым обвинением стандартной научной практики, существовавшей до нововведения[46].

Впрочем, предварительная регистрация — не серебряная пуля. Многие ученые, предварительно регистрирующие свои исследования, все равно не публикуют их (или не сообщают хотя бы о результатах) в установленный срок. Другие, несмотря на предварительную регистрацию, все равно вносят изменения в свой анализ после его начала[47]. В случае с клиническими испытаниями ученые не просто пренебрегают передовой практикой, но нарушают закон — в ходе одного расследования, проведенного журналом *Science* в 2020 году, обнаружилось, что о результатах свыше 55 % исследований ученые сообщили позже, чем обязаны были по реестру правительства США. Скажу лишь, что все должно было работать иначе: предварительная регистрация приносит пользу, только если соблюдается внесенный в реестр план. В области клинических испытаний нам необходимо ужесточить контроль за выполнением предписаний, а также наказания для "клинических нарушителей" (например, запрещать им в будущем получать правительственные гранты или публиковаться в определенных журналах в течение какого-то срока)[48]. В других областях, где нет установленной законом обязанности следовать зарегистрированным планам, нужно найти другие способы обеспечивать их выполнение учеными. В Великобритании, например, Национальному институту исследований в области здравоохранения и ухода удается добиться, чтобы почти все финансируемые им проекты оказывались опублико-

ванными, отчасти потому, что он удерживает 10 % от суммы гранта до тех пор, пока не появится отчет о проведенном исследовании[49].

Другой вариант — использовать еще более строгую версию предварительной регистрации. Ученые тогда подают на рецензирование саму регистрационную заявку, и если рецензенты ее одобряют и соглашаются, что дизайн исследования осмысленный, то журнал берет на себя обязательство опубликовать итоговую работу *независимо от того, какими окажутся результаты*. И только после этого ученые приступают к сбору данных[50]. Такой тип исследований (его называют *registered report*, "зарегистрированный отчет") не только кладет конец публикационному смещению, разрушая пагубную связь между статистической значимостью результатов и решением публиковаться, но еще и подрезает крылья *p*-хакингу, поскольку вы должны согласовывать свой анализ с рецензентами заранее и не можете просто изменить его *задним числом*, не объяснив предельно четко, что именно сделали. А самое главное, такой тип исследований сводит на нет многие из порочных стимулов, которые и приводят к предвзятостям и мошенничеству. Вы знаете, что опубликуете статью в любом случае, поэтому нет больше сильного давления, заставлявшего украшать свои результаты[51].

Предварительная регистрация — задумка удачная, если вынести за скобки практические аспекты контроля за соблюдением правил. Ведь в конечном итоге суть в *прозрачности:* не прячь свои планы или свой анализ от мира. Связанная с этим концепция "открытой науки" возникла намного раньше, чем кризис воспроизводимости, но может стать одним из мощнейших от него противоядий[52]. Открытая наука — это идея, согласно которой доступ к каждому этапу научного процесса нужно по мере сил делать свободным[53]. Идеальное исследование с точки зрения открытой науки имело бы со-

ответствующую веб-страницу, где можно было бы скачать все данные, весь статистический код, который использовался учеными для их анализа, и все материалы, что использовались для самого сбора данных[54]. И вместе с основной работой публиковались бы рецензии (даже если личность рецензентов не раскрывается) и предыдущие версии статьи, что позволяло бы читателю увидеть весь публикационный процесс[55].

Предоставление свободного доступа ко всем своим данным и методам воплощает мертоновскую норму коллективизма: вы позволяете другим ученым воспользоваться вашей работой, чтобы сделать их собственные исследования более эффективными. Еще оно служит норме организованного скептицизма: облегчает другим исследователям проверку вашей работы на наличие ошибок по небрежности, попытки воспроизвести ваши результаты и вообще более детальное изучение ваших открытий. Они смогут сделать все это гораздо проще, чем если бы использовали лишь итоговые таблицы и графики из вашей статьи, которые, как правило, только и прилагаются к журнальной публикации; причем отпадает необходимость писать вам напрямую и просить прислать данные, что, как мы знаем, срабатывает редко. Приверженность подходу открытой науки демонстрирует вашим коллегам: вам скрывать нечего, им не нужно слепо верить, что все было представлено точно.

Как следствие, если вам действительно *есть* что скрывать (вдруг вы подделали данные или результаты), открытая наука сильно осложняет вам жизнь. Как мы видели в третьей главе, создать фальшивый набор данных, который бы выглядел как настоящий, довольно трудно. Пока что расследователи, обнаружившие поддельные результаты, изучали, как правило, только сводные таблицы, представленные в статьях. Наличие доступа к полному исходному набору данных сильно упрощает подобные расследования. Но главное, открытые данные служат сдерживающим фактором против совершения мошенничества в принципе, ведь нужно быть наглейшим из наха-

лов, чтобы разместить подделку на общедоступном сайте[56]. Тот же принцип работает и в отношении *p*-хакинга, граничащего с мошенничеством, а также более невинных ошибок: вы позволяете другим ученым увидеть ваши данные и то, как вы их анализировали, а значит, зоркие коллеги могут заметить опечатки в таблицах, неверную статистику, неправдоподобные числа или неописанный анализ, от которого, возможно, зависит интерпретация ваших результатов.

Однако не любые данные можно разместить онлайн[57]. Например, размещение в интернете генетических данных участников исследования нарушит анонимность, которая служит непременным условием их участия, не говоря уже об их законном праве на неприкосновенность частной жизни. Предоставление общего доступа к некоторым другим видам данных также может быть опасно. В 2011 году, когда вирусологи опубликовали лишние подробности того, как они создали генетически модифицированную, более заразную версию вируса H5N1, вызывающего птичий грипп, разгорелись ожесточенные споры[58]. Это было важное исследование вирусов, но и редкая ситуация, когда последствия были бы ужасны, попади информация в руки людей с дурными намерениями. Исследователей попросили убрать из статьи некоторые методологические детали[59]. Впрочем, это исключения; в принципе, ничто не мешает подавляющему большинству ученых быть гораздо более открытыми в своей работе.

Есть еще один способ сделать науку открытой — расширить ее. Становится все очевиднее: чтобы ответить на серьезные научные вопросы, мы должны сделать еще один шаг по пути, который исторически начался с алхимиков-одиночек, ревностно охранявших свои секреты, а сегодня привел нас к тому, что ученые делятся результатами своей работы в международных журналах. Следующий шаг — так называемая командная наука, когда ученые из множества разных лабораторий и университетов сообща работают над более масштабными совместными проектами[60]. Такие совместные ис-

следования впервые были проведены в физике элементарных частиц, которая требует огромных команд ученых для выполнения тысяч невероятно сложных задач, связанных с экспериментами на ускорителях[61]. Еще совместные исследования становятся критически важными в современной генетике, где ученые объединяют наборы своих образцов в проектах международных консорциумов, что помогает достичь статистической мощности, необходимой для выявления слабеньких корреляций между десятками тысяч вариаций ДНК и человеческими признаками и болезнями[62]. Именно такой способ совместной работы выжег каленым железом ошибочный подход к генетике, основанный на "генах-кандидатах" (они обсуждались в пятой главе), и дал исследования, которые по крайней мере воспроизводятся[63].

Идея командной науки подхватывается во все большем количестве областей: например, попытка нескольких лабораторий повторить сотню известных психологических исследований, о чем мы говорили во второй главе, прекрасно иллюстрирует, как увеличение масштаба, а следовательно, и статистической мощности исследований может исправить ошибки прошлого. Как и в физике с генетикой, в настоящее время транснациональные совместные проекты ведутся в том числе в области нейронауки, эпидемиологии рака, психологии и трансляционных медицинских исследований[64]. Такие широкомасштабные проекты могут напрямую решать вопросы, связанные с воспроизводимостью в соответствующих областях, а поскольку результаты распространяются внутри большого сообщества, состоящего обычно из крайне категоричных ученых, теоретически эти проекты могут также служить своеобразной проверкой любого исследователя на предвзятость.

Открытая наука не ограничивается научным сообществом: финальный этап повышения прозрачности науки — открыть ее для широкой общественности. В конце концов, мертоновская норма коллективизма, которая подчеркивает совместное владение научными результатами, отражает важ-

ную мысль: огромная часть науки финансируется налогоплательщиками. Большинство журналов взимают со всех, у кого нет подписки, около тридцати пяти долларов за доступ к отдельной статье. Кажется проблематичным и даже недемократичным, что с представителей общественности, которая за счет своих налогов финансирует научные исследования, сдирают плату за доступ к полученным результатам. Разумеется, налогоплательщики финансируют много вещей, к которым не имеют доступа, что правильно: для них не раскрывают, например, британские секретные маршруты патрулирования подводных лодок, осуществляющих непрерывное ядерное сдерживание в открытом море. Однако этот вид информации — исключение, как и особые данные о вирусах в упомянутом выше случае. На сегодняшний день большинство правительств приняло законы о свободе информации, которые позволяют представителям общественности получать доступ к их документам и статистике; в США результаты работы всех государственных служащих автоматически считаются общественным достоянием[65]. От движения "Открытая наука" ответвляется движение "Открытый доступ", в рамках которого люди пытаются приблизить науку к такому состоянию свободной по умолчанию информации. Его влияние уже очевидно: многие журналы начали разрешать ученым вносить плату за свою публикацию, с тем чтобы все желающие могли скачивать эту статью бесплатно и бессрочно.

Раскошеливаться, чтобы ваши статьи стали общедоступными, необязательно, однако теперь финансирующие организации начинают *требовать*, чтобы оплаченные ими исследования публиковались в открытом доступе. Самая амбициозная стратегия под названием "План S" предложена ассоциацией *Science Europe*, представляющей финансирующие агентства европейских правительств. Идея в том, что с 2021 года все исследования, финансируемые членами этой ассоциации, должны публиковаться в журналах с полностью открытым доступом[66]. План S поддерживают правитель-

ственные советы по финансированию из более чем пятнадцати стран, а также крупные фонды, такие как *Wellcome Trust* и Фонд Билла и Мелинды Гейтс. Правилами предусматривается, что ученые, получившие гранты *Science Europe*, не могут даже подавать статьи на рассмотрение в издания без полного открытого доступа — а сюда входят и *Science* с *Nature*, и вообще 85% всех журналов[67]. Организации, финансирующие научные исследования, также заявляют, что будут действовать скоординированно, следя за тем, чтобы издатели не взимали с ученых слишком большую плату за публикацию в открытом доступе. Если журналы повысят взнос за открытый доступ, требуя больше некоторой разумной суммы для покрытия издательских расходов, фонды попросту введут лимит на оплачиваемые ими публикационные взносы. Это гарантирует, что огромная часть ученых будет вынуждена отказаться от таких журналов, что грозит им разорением. План S и в целом движение за открытый доступ демонстрируют, как коллективные действия со стороны финансирующих организаций могут привести к масштабным изменениям в исследовательской практике — и это должно вселить в нас надежду, что в будущем подобное сотрудничество поможет исправить и некоторые другие неполадки в науке.

Мы обсудили, какими способами предотвращать мошенничество, избегать недобросовестности и устранять предвзятость. А что насчет четвертого из наших научных изъянов — хайпа? Поскольку он касается того, как о результатах исследований сообщается, решение этой проблемы может быть связано с публикационной системой, которую мы только что обсуждали.

С первой главы этой книги мы считаем само собой разумеющейся издательскую модель, которая впервые появилась с журналом *Philosophical Transactions* в XVII веке: редактор при помощи советников-рецензентов решает, какие исследования

будут опубликованы в журнале, а затем обеспечивает их появление в печати. В не таком уж отдаленном прошлом, когда вся эта работа проводилась исключительно на бумаге, процесс рецензирования и публикации в журнале был невероятно трудоемким: переписка с рецензентами, предоставление им необходимых материалов, сбор и сопоставление их отзывов, затем редактура, корректура, печать и распространение самого бумажного издания. Понятно, что издатели брали плату за такую услугу. Однако с изобретением электронной почты и онлайн-журналов весь процесс значительно упростился. За что именно мы отдаем деньги в эпоху онлайн-публикаций, когда коммерческие издательства взимают непомерную плату за доступ к своим журналам?

Стоит подчеркнуть, *насколько* грабительская эта плата за подписку. В 2019 году лауреат Нобелевской премии Рэнди Шекман, биолог и редактор журнала, сравнил взносы, которые Калифорнийский университет выплачивает двум различным издательствам — некоммерческой Национальной академии наук и коммерческому издательству *Elsevier*[68]. За каждую скачанную научную статью университет платил Национальной академии наук 0,04 доллара. Плата издательству *Elsevier* за каждую статью составляла 1,06 доллара. *В двадцать шесть с половиной раз* больше. Исследователи со всего мира ссыпают миллионы в карман *Elsevier* — и миллиарды в казну коммерческих издательств в целом. Какие дополнительные услуги они получают за такие деньги? Неизвестно. Национальная академия наук предлагает в своих журналах фактически те же услуги, что и *Elsevier*, или даже лучше. Время и квалификация рецензентов — главный элемент всей научной системы, ведь эти люди обеспечивают критический анализ, необходимый для прогресса науки. Но рецензенты трудятся на добровольных началах, журналы не нанимают их на работу. Они бесплатная рабочая сила, поэтому никак не могут оправдывать большие взносы. Скорее *Elsevier* и другие коммерческие издательства занимаются тем, что экономисты на-

зывают поиском ренты, — получением большей прибыли без предоставления большей ценности[69]. В одном весьма критическом исследовании системы научных публикаций отмечается ее абсурдность:

> Это как если бы *New Yorker* или *Economist* потребовали, чтобы журналисты писали и редактировали работы друг друга бесплатно, и попросили правительство покрывать расходы. Устройство этой системы обычно повергает сторонних наблюдателей в некое ошеломленное неверие. <… > В отчете *Deutsche Bank* за 2005 год она названа “странной” системой “тройной оплаты”, когда “государство финансирует большинство исследований, выплачивает зарплату большинству тех, кто проверяет качество исследований, а затем покупает бóльшую часть опубликованной продукции”[70].

Если вспомнить, что взносы, которые университеты платят компаниям вроде *Elsevier*, занимающимся поиском ренты, в значительной степени поступают из государственного бюджета, то все это начинает казаться аморальным: разве университеты не должны как-то эффективнее расходовать деньги налогоплательщиков? Дело даже не в том, что нужно препятствовать погоне за прибылью или участию коммерческих издательств в научной деятельности. Иррациональная система престижа, сложившаяся вокруг журнальной публикационной системы, олицетворяет крах рынка, задавливая необходимую конкуренцию, которая могла бы снизить цены. Трудно утверждать, что издатели журналов добавляют статьям такую уж большую ценность: часто они служат лишь посредниками между авторами и добровольными рецензентами, практически не занимаясь редактированием или другой работой, кроме обычной верстки[71].

В журнальной публикационной системе могут произойти большие изменения, и мы становимся свидетелями одного из них прямо сейчас. Я говорю о росте числа *преприн-*

тов. Препринт — это черновик научной статьи, размещенный онлайн в бесплатном хранилище, чтобы любой желающий мог с ним ознакомиться. Предполагается, что другие ученые прочитают препринт и прокомментируют исследование, а автор внесет необходимые поправки до того, как подаст статью в журнал для официальной публикации, — новая вариация на тему социальных процессов в науке. Экономисты и физики пользуются схемой с препринтами уже десятки лет, но в последние годы препринты захватили и другие дисциплины, например биологию, медицину и психологию[72]. В такой области, как генетика, где новые методы, данные и результаты появляются с головокружительной скоростью, препринты позволяют ученым быть в курсе всех последних достижений без длительного ожидания формальных публикаций. Еще благодаря препринтам ошибки, неудачные идеи и плохие статьи могут быть немедленно найдены и подвергнуты критике любым представителем научного сообщества (или внешнего мира), а не двумя-тремя людьми, которые будут участвовать в обычном процессе рецензирования. Такая критика — как и любые положительные комментарии — обычно тоже публикуется онлайн, что делает рецензирование более неформальным, но в то же время более прозрачным, чем в стандартной модели[73].

Препринты уже ускорили темп своего распространения и повысили открытость научных исследований, а еще, будем надеяться, уменьшили необходимость откладывать "провалившиеся" исследования в "картотечный ящик", поскольку им не требуется принятие журналом. Но как это решает проблему хайпа? Как мы уже видели, одна из основных причин неоправданного преувеличения — необходимость убеждать рецензентов и редакторов в том, что статья достойна опубликования: ученые раздувают последствия своих открытий и раскручивают отрицательные результаты в попытке заполучить заветные публикации. Роль журналов заключается не только в распространении исследований в мире — они

решают, что будет опубликовано. Что, если бы мы полностью разделили эти две роли?[74] Одно радикальное предложение таково: после завершения исследования ученые пишут препринт и загружают его в онлайн-хранилище. Затем они просят, чтобы его взяла на рассмотрение служба рецензирования — организация нового типа, которая набирает рецензентов обычным способом, но не имеет отношения к какому-либо научному журналу[75]. Служба рецензирования рассматривает препринт и дает ему оценку. Авторы при желании могут его доработать (собрать дополнительные данные или что еще потребуется), а потом повторно представить улучшенную версию для получения новой оценки — более высокой, как они надеются. Редакторы журналов, чья роль в такой системе больше похожа на кураторскую, чем на охранительную, внимательно изучают все оцененные препринты и выбирают те, которые хотят включить в свой журнал.

В такой модели все публикуется в виде препринтов, а журналы становятся усилительными устройствами для лучших или наиболее актуальных исследований — как газеты, что печатают самые важные истории от информационных агентств вроде *Reuters* и *Agence France-Presse*. Журналами, на которые стоило бы подписаться, будут те, что обеспечивают наилучшую дополнительную ценность, просеивая тысячи новых препринтов в поисках наиболее достойных вашего внимания[76]. Таким образом, конкуренция, подстегивающая написание более высококачественных статей, все равно будет присутствовать: ученые хотят, чтобы их работы оказались выбраны самыми известными редакторами-кураторами, и соперничают по тому, *где* публикуются их статьи. А вот решение, публиковать ли их *вообще*, будет уже принято[77].

Эта идея — не единственное верное решение проблем науки, на деле подобная модель может оказаться не лучше, чем существующее положение вещей, или даже содержать неведомые неизвестные, которые ухудшат ситуацию. Попробовать, безусловно, стоит[78]. Однако нужно быть осторож-

ными с этими новыми публикационными моделями. Как нам контролировать качество науки в мире, где любой может запросто опубликовать свои результаты онлайн без какого-либо рецензирования?[79]

Поучителен пример с препринтом 2016 года, написанным гарвардским экономистом Роландом Фрайером на одну из самых спорных тем, касающихся современной американской действительности, — о применении полицией силы против чернокожих людей[80]. Анализ Фрайера по большей части основывался на подробных данных об арестах, произведенных полицией в Хьюстоне, штат Техас. Результаты оказались парадоксальными: хотя против чернокожих и латиноамериканцев более чем на 50 % чаще, чем против белокожих, во время задержаний применялась сила с несмертельным исходом, в отношении силы со смертельным исходом все было как раз наоборот — “шансы чернокожих быть застреленными полицией на 23,8 % *меньше* по сравнению с белокожими”[81]. Препринт привлек внимание общественности благодаря передовице в газете *The New York Times* (“Удивительные новые данные свидетельствуют о предвзятости при применении полицией силы, но не при стрельбе”), где цитировались слова Фрайера “самый неожиданный результат в моей карьере”[82]. Это быстро подхватили обозреватели-консерваторы, недовольные движением *Black Lives Matter* (“Жизни чернокожих имеют значение”) против насилия со стороны полиции. Один из них написал, что исследование Фрайера “доказывает”: движение “построено на лжи”[83].

Значение 23,8 % получилось у Фрайера просто на основании сырых данных. В исследовании были далеко не только они, но при освещении в средствах массовой информации остальное, к сожалению, затерялось, особенно после того, как Фрайер сам обратил внимание журналистов на это конкретное число. После внесения поправок на другие детали каждого случая, такие как наличие оружия у задержанного, результат поменялся на противоположный: в сопоставимых

ситуациях вероятность быть застреленными у чернокожих превышала таковую у белокожих. Это различие между необработанными и скорректированными результатами согласуется с расовыми предрассудками в рядах полиции: именно такой эффект можно было бы наблюдать, если бы полицейские просто арестовывали больше чернокожих вообще, независимо от того, представляли ли те угрозу, и чаще застреливали бы чернокожих арестованных при обострении ситуации. Иными словами, число чернокожих задержанных, которых застрелили, было размыто, поскольку сырые данные включали в себя множество не представлявших угрозы чернокожих, которых арестовали без всякой причины[84]. Однако не стоит вкладывать сюда особый смысл и придавать всему этому слишком большое значение: выборка Фрайера чересчур мала. Различие в 23,8 % даже не является статистически значимым[85]. Ограничься Фрайер подчеркиванием только статистически значимых результатов, всей этой полемики вообще бы не возникло.

Если бы журналисты *The New York Times* дождались опубликованной версии статьи, которая вышла в апреле 2019 года, они обнаружили бы подачу информации под несколько иным углом: в аннотации выражение “на 23,8 % меньше” было опущено и просто отмечалось, что различий по показателю летальности между чернокожими и белокожими людьми не найдено[86]. Хотя и это любопытно, такая формулировка была довольно большим отступлением от первоначального утверждения Фрайера. Итак, мораль этой истории в том, что надо остерегаться ученых с препринтами? Поскольку препринты можно разместить в интернете без какого-либо надзора, нам точно нужно относиться к ним особо скептично, а ученые должны иметь интеллектуальную скромность, чтобы не обнародовать свою работу до того, как ее хотя бы просмотрят коллеги[87].

По мере изменения научной экосистемы журналисты будут все четче понимать, что у научной публикации существуют различные “стадии” и следует с особой осторожно-

стью относиться к работам, находящимся на стадиях ранних. Вскоре после начала пандемии коронавируса (COVID-19) в 2020 году на крупнейшем сервере препринтов по биологии появились работы, вызвавшие широкое обсуждение происхождения и действия вируса. Некоторые из этих препринтов, явно низкого качества, были написаны в спешном порядке, чтобы извлечь выгоду из раздуваемой в средствах массовой информации истерии вокруг пандемии. Другие содержали формулировки, словно бы разжигающие — непреднамеренно или наоборот — теорию заговора: что вирус якобы специально разработали как биологическое оружие. В ответ на это (и в дополнение к своему постоянному предупреждению, что размещаемые ими препринты не рецензируются) архив добавил в верхнюю часть каждой страницы следующую запись:

> Напоминание: это предварительные отчеты, не проходившие рецензирование. Они не должны рассматриваться как окончательные, направлять клиническую практику и поведение в отношении здоровья или освещаться в новостях как подтвержденная информация[88].

Однако на протяжении всей этой книги мы видели, что рецензирование — не гарантия научного качества. Вспомните исследование вакцины MMR Эндрю Уэйкфилда — возможно, самое пагубное научное мошенничество всех времен. Оно не просто прошло рецензирование, оно было отрецензировано и принято к публикации в издании *The Lancet*, одном из самых престижных медицинских журналов мира. Хотя препринты могут выступать переносчиками ошибочной информации, вряд ли нам удастся разработать систему, исключающую любые ошибки. Необходимо сравнить отрицательные стороны препринтов с положительными, ведь они обеспечивают открытость, прозрачность и оперативность. И действительно, вирусологам и эпидемиологам, отвечающим на кри-

зис, вызванный коронавирусом, революция в области препринтов принесла волну новых данных, что существенно ускоряет науку, формируя исследовательскую культуру, совершенно отличную от той, что складывалась во время предыдущих вспышек заболеваний. Отсутствие необходимости ожидать формального рецензирования, возможность мгновенно комментировать черновики новых статей и делиться важными отрицательными результатами, которые обычно не преодолевают публикационное смещение, дали нам — несмотря на редкие неверные заявления — научную литературу, на месяцы, а то и на годы опережающую то свое состояние, в коем она пребывала бы в противном случае. Хотя еще неясно, спасет ли нас эта ускоренная наука от последствий самой страшной пандемии за последние сто лет, которая на момент написания этих строк уносит тысячи жизней и схлопывает мировую экономику, урок для науки очевиден: в состязании между старой исключительно журнальной схемой и той, что дополнена препринтами, система с препринтами выигрывает с большим отрывом[89].

Приведенные выше идеи обнадеживают, однако направлены они в основном на устранение *симптомов* современных недугов науки, а не их причин. Если игнорировать первопричины, то всех потенциальных реформаторов науки ожидает сизифов труд — постоянное восполнение ущерба, наносимого каждым новым поколением ученых, что стремятся украсить свои резюме захватывающими, “новаторскими” результатами. В предыдущей главе мы видели, как это стремление вызывается давлением извне. Ученым нужны работа и гранты, и их нескончаемая погоня за результатами, обещающими публикации в журналах с высоким импакт-фактором, во многом является погоней именно за статьями. Обратите внимание, например, сколь многие университеты заставляют своих ученых для получения штатной должности

подписывать соглашение о количестве работ, которые нужно будет выдавать, и об импакт-факторах журналов, где те должны будут публиковаться[90].

Странно, впрочем, что ученые, уже состоящие в штате и руководящие прекрасно финансируемыми лабораториями, продолжают регулярно заниматься всякими дурными делами, которые мы обсуждали в этой книге. Порочные стимулы так глубоко укоренились, что создали самоподдерживающуюся систему. Годы явной и неявной дрессировки любой ценой добывать публикации и цитирования оставляют свой след на молодых ученых, формируя новые нормы, привычки и способы мышления, которые трудно вытравить, даже когда стабильная работа уже обеспечена. И, как мы обсуждали в предыдущей главе, такая система создает давление отбора, когда на плаву удерживаются только те научные сотрудники, кто по своей природе хорош в этой игре. Если вы из тех людей, чье ненасытное эго заставляет вас делать все, чтобы превзойти по индексу Хирша своих коллег, сейчас академическая система встречает вас с распростертыми объятиями.

Самоподдерживающиеся системы особенно неподатливы. Да, изменение системы академических стимулов и ее извращенных требований непрерывно публиковаться и получать все больше грантов критически важно для исцеления науки. Три главных игрока, ответственных за эти перемены: университеты, финансирующие организации и журналы. Они обладают необходимой властью, чтобы спустить давление, нагнетаемое лозунгом “Публикуйся или погибни”, и мы сейчас увидим, как именно они могут это сделать. Но если на том мы и остановимся, то не учтем, что многие ученые никакой другой игры до сих пор и не знают. Чтобы переломить ситуацию, нам необходимо объединить изменения, нисходящие от крупных игроков, с силами, восходящими от самого научного сообщества.

Чтобы самим ученым легче было согласиться с необходимыми реформами, мы должны признать, что на мертоновской

норме бескорыстности, как бы ни была она достойна восхищения, далеко не уедешь. На практике личное вознаграждение является и всегда являлось важным стимулом к научному прогрессу[91]. Если мы хотим побуждать исследователей теснее сотрудничать друг с другом и свободно делиться своими данными в интернете, нам нужно гарантировать, что они получают то вознаграждение, какого заслуживают.

Это потребует серьезной работы и существенных изменений в том, как мы оцениваем усилия исследователей. Но еще у нас будет прекрасная возможность заменить нынешнюю непрозрачную и зачастую несправедливую систему описания научного сотрудничества. В настоящее время в большинстве научных областей для статей с многими авторами принято указывать первым имя ученого, выполнившего основную часть работы, а последним — имя более опытного научного сотрудника, курировавшего проект[92]. Все остальные — кто, скажем, помогал продумать план исследования, организовать лабораторию, набрать участников, собрать материалы и данные или провести статистический анализ — перечисляются где-то в середине. Как способ отдать должное и воздать по заслугам вариант далеко не идеальный. С одной стороны, многие из тех, кто указан в списке "авторов", могут не иметь практически или вовсе никакого отношения к самой статье; с другой же — реальный вклад многочисленных авторов из середины остается непризнанным, поскольку университеты при распределении должностей и зарплат в основном ориентируются на публикации, где ученый стоит первым либо последним автором.

Масштабные, коллективные действия в рамках "командной науки" проблему только усугубят: они часто приводят к появлению работ, чьи списки авторов насчитывают сотни, а то и тысячи имен[93]. Обмен данными, необходимый для открытой науки, также внушает опасения по поводу невознаграждаемых усилий. Что, если это породит новый класс ученых — "исследователей-паразитов", как их назвали в редакционной

статье в издании *The New England Journal of Medicine*, — которые сами никогда не собирают данные, а просто полагаются на результаты анализа тех данных, что кропотливо собирались другими учеными?[94] Наша нынешняя система расплывчатых неписаных правил и индивидуальных индексов цитирования малопригодна для решения этой проблемы.

Требуется, таким образом, новый способ распределения лавров и ответственности в науке — система, при которой ученые вознаграждаются за свой вклад, а не просто за то, что их имена указаны в статье[95]. Для достижения этой цели необходимо, чтобы первые шаги сделали университеты. При принятии решений о приеме на работу и зачислении в штат они должны учитывать "социально ответственную научную позицию" ученого в дополнение к таким показателям — или даже вместо них, — как индекс Хирша[96]. Они должны признавать не только публикации, но и сложность создания международных коллабораций, трудоемкость сбора данных и обмена ими, честность при обнародовании каждого исследования, положительные ли оно дало результаты или отрицательные, и малопривлекательный, но необходимый процесс проведения исследований-повторений. Иными словами, университеты должны вознаграждать ученых за усилия, направленные на создание более открытой и прозрачной научной литературы, расширяя диапазон того, что считается целесообразным[97]. Они должны явно ценить качество, а не количество и больше внимания уделять методам ("Проводит ли этот человек скрупулезную работу, которая дает надежные научные данные?"), а не результатам ("Получает ли этот человек p-значения ниже 0,05?"). И еще университеты должны рассматривать гранты, по выражению психолога Скотта Лилиенфельда, "как средство достижения цели, а не как самоцель"[98].

Финансирующие организации должны последовать этому примеру: вместо того чтобы основывать свои решения на количестве предыдущих работ исследователя или громкости его

заявлений, им следует ввести критерии, которые потребуют от ученых делать на выделенные деньги нечто большее, чем просто публиковать статьи в журналах с высоким импакт-фактором. Сколько данных будет получено в ходе предлагаемого исследования? До какой степени они будут представлены обществу? Финансирующие организации могли бы также следовать инициативам, рассмотренным выше, и удерживать деньги, если ученые способствуют публикационному смещению, сообщая не обо всех своих результатах. Как показывает план S, когда финансирующие организации объединяются для продвижения принципов открытой науки, они являют собой грозную силу.

Можно и нужно претворять в жизнь и гораздо более радикальные изменения. Чтобы остановить гонку вооружений между учеными, конкурирующими за финансирование посредством все более раздутых заявлений, предлагаются различные методы. Одна идея, которую какие-то финансирующие организации уже применяют, заключается в том, чтобы преимущественно финансировать *ученых*, а не их конкретные проекты. Иначе говоря, тем ученым, которые считаются довольно изобретательными, вы можете предоставлять достаточно денег на запланированные ими исследования без ограничений и надеяться, что возросшая свобода дарует новые и интересные достижения. Это дало бы исследователям передышку от постоянной борьбы за финансирование и позволило бы им сосредоточиться на более долгосрочных и добротных проектах, вместо того чтобы вынужденно рассылать по журналам короткие, возможно нарезанные на кусочки, статьи ради улучшения своего резюме перед следующей заявкой на грант[99].

Но это не очень-то подходит для более масштабных идей, когда планировать тематику исследований и закладывать в бюджет траты на соответствующий персонал и оборудование нужно заблаговременно — что может привести к несправедливости по отношению к менее известным ученым, у которых идеи хороши, но нет связей и влияния, необ-

ходимых для отбора на долгосрочное финансирование. Как лучше распределять деньги в мире, где мы все еще финансируем проекты? Особенно заманчивой выглядит идея создать шорт-лист заявок на гранты, качество которых превышает определенный уровень, а затем распределять финансирование жеребьевкой. Учитывая, что научная система считается меритократической, звучит это странновато. Однако, по мнению сторонников жеребьевки, нынешняя система настолько плоха в распределении средств, что "это уже, по сути, лотерея, только без преимуществ случайности"[100]. Проведенный в 2016 году анализ показал, что оценки, выставленные рецензентами заявкам на гранты Национальных институтов здравоохранения США, практически не были связаны с качеством исследований, проведенных на эти гранты (оно измерялось количеством ссылок на итоговую статью)[101]. Если так, то значительная часть времени, уделяемого учеными на доработку своих заявок на гранты, чтобы сделать их интереснее и убедительнее, тратится впустую. И действительно, согласно одному подсчету, "ценность науки, которую исследователи упускают при подготовке заявок на гранты, примерно равна или превышает ценность науки, поддерживаемой программой финансирования"[102].

Вероятно, невозможно справедливо провести отбор заявок на гранты из россыпи высококачественных претендентов. Если отдать это на волю случая, рецензентам не нужно будет тратить время на попытки одолеть невыполнимое, а также это позволит избежать предвзятости по отношению к определенным типам ученых из-за их стажа, гендера или других характеристик[103]. Знай ученые, что их проектная заявка будет участвовать в лотерее при условии, что она не ниже определенного уровня по качеству — а планку можно установить достаточно высоко и учитывать такие критерии, как "следование принципам открытой науки", — они бы испытывали куда меньше давления, заставляющего преувеличивать важность своей работы[104].

Журналы тоже должны принять новые стандарты, способствующие открытости и воспроизводимости. Было бы прекрасно, если бы для начала они стали приглашать ученых представлять на рассмотрение исследования-повторения, предварительно регистрировать свои планы и сопровождать статьи соответствующими наборами данных[105]. Это увеличивает количество способов, какими ученые могут заинтересовать своей статьей редакторов журналов вне узкой направленности на положительные результаты, и означает, что исследователи с отрицательными результатами не потерпят неудачу на первом же этапе[106]. Еще журналы могли бы ввести процедуру сканирования статей на предмет обычных ошибок или нанять аналитиков, которые бы проводили выборочные проверки данных на достоверность с использованием таких методов, как тест GRIM[107]. Журналы уж точно должны побуждать авторов представлять свои открытия со смирением: подобно тому как они требуют от ученых сообщать о конфликтах интересов, они могли бы требовать от них писать раздел “Ограничения исследования”, который сопровождал бы в небольшой рамке итоговую публикацию[108]. Они даже могли бы попытаться отвадить ученых от нарезки статей, сделав подачу публикаций на рассмотрение более затратной (по времени, усилиям или деньгам)[109].

Все это университеты, финансирующие организации и журналы должны или могли бы сделать. Но что им с того? В конце концов, университеты хотят обеспечить себе статус на все более безумном рынке, который оценивает их по количеству статей, публикуемых работающими на факультетах учеными в журналах с высоким импакт-фактором; финансирующие организации хотят, чтобы все видели, как они поддерживают самую престижную науку; журналы хотят наблюдать, как с каждым годом растет их импакт-фактор. Однако настало время перемен, и тому есть две причины.

Первая — сам развивающийся кризис воспроизводимости. Десятилетиями, даже веками, многие ученые питали серь-

езные подозрения насчет достоверности результатов и насчет порождающей их системы. Но лишь в последние несколько лет накопились горы надежных данных, подтверждающих эти подозрения. Теперь во многих областях мы можем поместить отдельные громкие случаи ошибок в науке в более широкий контекст системных проблем и указать на убедительные метанаучные данные, подтверждающие частные случаи. Иными словами, только недавно мы обрели данные, которые закладывают реальную основу для исцеления науки.

Ни один университет не желает славиться тем, что работы его ученых не воспроизводятся или что его сотрудники замешаны в расследованиях мошенничества. Не хотят такого и финансирующие организации: они понимают, что если финансируют не высококачественные, воспроизводимые исследования, то тратят деньги впустую, а в конечном итоге будут выглядеть дураками. То же относится и к журналам: последнее, чего они сами или их издатели хотят, — стать посмешищем, как это произошло с некоторыми изданиями, приобретшими дурную славу из-за публикуемых в них сомнительных исследований[110]. В эту передрягу мы попали отчасти именно из-за репутационных соображений, но, если скандалы и изъяны станут достаточно известными, те же самые соображения помогут нам из нее и выпутаться.

Вообще беспокойство из-за кризиса воспроизводимости среди самих ученых уже вызвало широкую поддержку реформирования науки. Ученые продемонстрировали готовность *массово* отказаться от стандартной публикационной системы, о чем свидетельствует их энтузиазм в отношении препринтов. Научное сообщество увидело преимущества препринтов в скорости и прозрачности — и решилось, а за ним последовали и журналы, изменив свою политику так, чтобы статьи, вышедшие поначалу в виде препринтов, все еще принимались к публикации. Аналогичный спрос на открытый доступ, открытые данные и предварительную регистрацию тоже влияет на политику журналов.

Сцепка между требованиями снизу и правилами сверху может привести к возникновению положительных обратных связей — добродетельному кругу, создающему полезные самоподдерживающиеся нормы, а не порочные. Например, "Британская сеть воспроизводимости" (UKRN, *UK Reproducibility Network*), созданная научно-педагогическими работниками, обеспокоенными кризисом воспроизводимости, и представленная общественными организациями в университетах, в настоящее время обсуждает с этими университетами, как изменить практику найма сотрудников, чтобы поощрять открытость и прозрачность[111]. Такие усилия могут послужить противоядием к "естественному отбору плохой науки", который мы обсуждали в предыдущей главе. Поощрение ученых, ценящих открытость и прозрачность, может создать свой собственный добродетельный круг: неуклонное увеличение числа специалистов, практикующих открытую науку, будет способствовать дальнейшим изменениям снизу и побуждать других исследователей присоединяться к движению и связанным с ним реформам. Это породит еще больше метанаучных работ, помогающих понять, какие виды исследований воспроизводятся, какие академические структуры содействуют проведению надежных исследований и к каким результатам нам в общем и целом следует относиться с подозрением[112].

Одобрение и уважение коллег прекрасно мотивируют, так что не стоит недооценивать и силу стыда. Страх быть осмеянным в интернете за глупые ошибки, проявление явной предвзятости или чрезмерное раздувание важности своих исследований должен служить веским основанием для перепроверки числовых данных и удержания собственных заявлений в рамках реальности[113]. Но не нужно концентрироваться исключительно на негативных предпосылках для изменений. Вторая важная причина, по которой сейчас самое время преобразовать научную систему, — это технологии.

Попросту говоря, заниматься открытой наукой никогда еще не было так легко. Алгоритмы теперь умеют проверять

статьи на наличие ошибок; препринты и предварительная регистрация доступны мгновенно; даже очень большими наборами данных можно делиться такими способами, какие еще несколько лет назад казались немыслимыми; вклад авторов в работу над статьями и в функционирование научного сообщества в целом можно отслеживать в мельчайших деталях; вся цепочка от сбора данных до публикации может быть представлена на всеобщее обозрение. Брайан Носек, директор Центра открытой науки, обращает на это наше внимание: прежде чем браться за глубоко укоренившиеся нормы, стимулы и стратегии, нужно сделать первый, обманчиво простой шаг к культурным переменам — создать условия для того, чтобы люди легко восприняли ваши новые идеи[114]. Очень многие ученые хотели бы повысить качество своей работы, но их сдерживает опасение, что такие изменения могут потребовать неподъемных усилий. Новые и постоянно развивающиеся технологические решения способны устранить это опасение и позволить все большему числу исследователей присоединяться к движению открытой науки.

Если недостаточно просто сделать этот переход возможным и несложным, то мы вправе апеллировать к личным интересам ученых. Специалист по онкобиологии Флориан Марковец в своей статье о том, как ученым следует пользоваться преимуществами новых автоматизированных инструментов, чтобы делать прозрачными связи между своими данными, их анализом и публикациями, приводит “пять корыстных причин работать воспроизводимо”:

1. Открытость и прозрачность данных помогают вам с соавторами *выявлять ошибки*, способные подорвать ваши результаты, и не дают исследованию превратиться в катастрофу вроде случая с опечаткой в электронной таблице у Рейнхарт и Рогоффа.
2. Новые автоматизированные методы *облегчают написание* статьи.

3. Если любой желающий может посмотреть, как вы анализировали свои данные, то легче *убедить рецензентов*, что все у вас сделано правильно (а если доступны данные, рецензенты даже могут попробовать провести ваш анализ самостоятельно).
4. Открытое документирование каждого этапа анализа поможет вам *продолжить свою работу* несколько месяцев спустя: "вам будущему" не придется полагаться на воспоминания о том, что делали "вы прошлый".
5. Делая все доступным, вы показываете научному сообществу, что работали добросовестно и скрывать вам нечего, — иными словами, вы *создаете себе репутацию* честного исследователя[115].

Мы четко видим проблему. Решения осуществимы. Все, что нужно для исцеления науки, — правильно мотивировать людей.

Запись третьей симфонии Хенрика Гурецкого, "Симфонии печальных песен", разошлась тиражом более миллиона экземпляров, что абсолютно неслыханно для современной классической музыки[116]. Отчасти такая популярность объясняется простотой симфонии: она нетороплива, почти как сопровождение кинофильма, без резкой атональности, присущей предыдущим произведениям Гурецкого. Несмотря на итоговые невероятные продажи, премьера 1977 года определенно понравилась не всем слушателям. Когда финальная часть отзвучала двадцать одним повторением аккорда ля мажор, Гурецкий услышал, как "известный французский музыкант", сидевший в первом ряду (большинство полагает, что это был вспыльчивый авангардист Пьер Булез), прошипел: "*Merde!*"*[117] Хотя сегодня третья симфония Гурецкого при-

* Дерьмо! *(франц.)*

знана шедевром современной музыки, для музыкантов-революционеров она просто не была достаточно новаторской или экспериментальной.

Ученые мыслят скорее как Булез, чем как Гурецкий. Они слишком возвеличили новизну в науке, породив извращенную неофилию, когда каждое исследование непременно должно стать грандиозным прорывом, в корне меняющим наше представление о мире, а возвещать о великом открытии должен, видимо, ученый в белом лабораторном халате, взволнованно влетающий в комнату и размахивающий листками бумаги, как это происходит в фильмах. Ученые хотят, чтобы их открытия выглядели так, словно заслуживают вскриков “Эврика!”, поэтому они анализируют, описывают и публикуют свои исследования соответствующим образом. И хотя неожиданные прорывы время от времени действительно случаются, преимущественно наука развивается постепенно и кумулятивно, потихоньку продвигаясь в сторону пристрелочных теорий, а не резко подскакивая к окончательным истинам[118]. Бóльшая часть науки, если совсем уж честно, довольно скучна.

Если мы снизим градус хайпа вокруг новых результатов и начнем смиреннее относиться к тому, что знаем, наука может сделаться скучнее — однако скучное, но достоверное должно побеждать захватывающее, но неосновательное практически при любых обстоятельствах. Как публикация большего количества отрицательных результатов и исследований-повторений — это более надежный способ увеличивать наше общее знание, так и осознание предварительности и неопределенности самой природы исследований — это в долгосрочной перспективе лучший способ в полной мере оценить науку. Давайте работать над тем, чтобы не поддаваться неофилии, сорочьей падкости на блестящие результаты исследований и вместо этого ценить результаты, которые убедительны, даже если прямо сейчас они и не столь захватывающи. Иными словами, давайте СНОВА СДЕЛАЕМ НАУКУ СКУЧНОЙ[119].

Но не *слишком* скучной. Наука только *преимущественно* развивается постепенно; мы не хотим, чтобы при изменении стимулов маятник отклонился слишком далеко в противоположную сторону, отбивая у ученых желание дерзать, идти на риск с новыми смелыми, неожиданными идеями и заниматься изысканиями, которые нет-нет да и приводят к крупным открытиям. Фокус будет состоять в том, чтобы найти оптимальный баланс между нашим новым, “пошаговым” отношением к знанию и пониманием, что иногда прожектерские исследования эксцентричных персонажей в духе Булеза действительно дают огромную отдачу[120]. Эти точки зрения примирить легче, чем кажется: не только нельзя оспаривать тот факт, что унылая сосредоточенность на публикациях и цитированиях сдерживает необычные поисковые исследования, которые могли бы привести к революционным достижениям[121], но и описанные в этой главе виды реформ сделают новые результаты более осмысленными, разграничивая игру случая (или человеческой предвзятости и подачи под нужным углом) и подлинные волнующие зацепки, которые могут подсказать нам *следующее значительное открытие*.

Вообще в научном поиске есть определенное благородство — когда ученый следует мертоновским принципам универсализма, коллективизма, бескорыстности и организованного скептицизма в искреннем стремлении к истине и знаниям о мире. Но все эти добродетели подвергаются коррозии в системе, поощряющей возмутительное раздувание результатов, ревностную охрану данных, халатное срезание углов, бесстыдную погоню за престижем и безобразное мошенничество. Если нам удастся обучить новое поколение исследователей придерживаться мертоновских норм и в то же время сдерживать поток стимулов, который толкает в противоположном направлении, мы, возможно, сумеем спасти науку от себя самой.

Разумеется, потребуется некоторое время, чтобы убедить научное сообщество попробовать претворить в жизнь намеченные в этой главе идеи. Нам нельзя быть в таком вопросе догматиками: это ненаучно само по себе, да еще и разные области науки нуждаются в разных реформах — не существует единого, универсального решения[122]. Мы должны действовать осторожно, экспериментируя с новыми методами работы и собирая о них информацию, а не крошить существующую систему, навязывая реформы неким постановлением.

В идеале метанаучные доказательства, приведенные в этой книге, должны убедить почти любого читателя в том, что с наукой все совсем не в порядке и назрела острая необходимость в изменениях. Но даже если вы считаете, что говорить о "кризисе" — слишком большое преувеличение, у меня в арсенале остался еще один, последний аргумент[123]. Вот он: все реформы, которые мы обсуждали в этой главе, принесли бы науке пользу, *даже если бы кризиса воспроизводимости и не было*. Тут мне вспоминается классическая карикатура Джоэла Петта на тему изменения климата из газеты *Lexington Herald-Leader*:



Переиначу текст с этой карикатуры (надеюсь, Петт меня простит), чтобы дать ответ сомневающимся:

> Открытость. Прозрачность. Улучшенная статистика. Предварительная регистрация. Автоматизированная проверка ошибок. Хитроумные методы отлова мошенников. Препринты. Более эффективная практика найма. Новая культура смирения. И прочее и прочее. А вдруг кризис воспроизводимости — это большая мистификация и мы напрасно создаем лучшую науку?

Эпилог

Не надо лгать — и победите черта*.
Уильям Шекспир *“Король Генрих IV (Часть первая)”, акт 3, сцена 1*

Пока я писал эту книгу, астрофизики получили первую “фотографию” черной дыры[1]. Медицинские генетики объявили, что семеро детей с тяжелым иммунодефицитом, которые вынуждены были жить в изоляции, чтобы не подхватить какую-нибудь распространенную — но для них смертельную — инфекцию, благодаря генной терапии, похоже, излечились; а генная терапия при муковисцидозе показала результаты, позволяющие надеяться, что она будет помогать 90 % людей, страдающих этим заболеванием[2]. Специалисты в области здравоохранения отчитались о том, что у ВИЧ-положительных гомосексуалов, принимающих новейшие антиретровирусные препараты, вероятность передачи вируса сексуальному партнеру “фактически равна нулю”[3]. Инженеры телепортировали информацию внутри алмаза с помощью квантовой запутанности[4]. Ученые ввели наночастицы в глаза мыши, наделив их способностью видеть в инфракрасном диапазоне[5]. Все это настоящие чудеса, и их появление в ряду научных и медицинских открытий постоянно напоминает нам, что наука — одно из величайших достижений человечества.

По крайней мере, это должны быть чудеса. Нам следует гордиться ими. Но у знания обо всех изъянах науки, обсу-

* Перевод Б. Пастернака.

ждавшихся в этой книге, есть удручающее следствие: мы можем начать подозрительно относиться к любому и каждому новому результату, зная, что поток научных открытий далеко не чист. Примерно в то же время, когда были сделаны достижения, описанные выше, Служба по обеспечению добросовестности в научной практике США заключила, что медицинский исследователь из Университета Дьюка фальсифицировал данные тридцати девяти опубликованных статей и более чем шестидесяти грантов на общую сумму свыше двухсот миллионов долларов[6]. Обнаружилось, что профессор генетики так "беспутно" руководил своей лабораторией в Университетском колледже Лондона, что из нее вышли предположительно десятки жульнических статей, однако же он отказался уволиться или взять на себя какую-либо ответственность[7]. В нашумевшем исследовании по психологии, опубликованном в *Science*, заявлялось, что у консерваторов физиологические реакции (например, на резкие звуки) сильнее, хотя оно не воспроизвелось на куда большей выборке, а журнал — что более или менее точно повторяет историю нашей провалившейся попытки воспроизвести экстрасенсорную работу Бема — с порога отмел повторную статью[8]. Новый алгоритм, борющийся с плагиатом, обнаружил свыше семидесяти тысяч статей на русском языке, которые были опубликованы как минимум дважды, а какие-то из них появлялись во многих разных журналах — числом до семнадцати[9]. Один из самых высокоцитируемых исследователей в мире, американский биофизик, был изгнан из редакционной коллегии биологического журнала, когда обнаружилось, что он регулярно принуждал авторов, чьи статьи редактировал, ссылаться на его собственные публикации (иногда на более чем пятьдесят за раз), что очевидным образом обеспечивало ему резкий скачок числа цитирований[10].

Мы не можем просто дивиться новым научным достижениям, ведь нам известно, что надежные, воспроизводимые результаты идут рука об руку с тьмой ошибочных, предвзятых,

вводящих в заблуждение и фальсифицированных исследований. Нефролог Драммонд Ренни в 1986 году прекрасно выразил эту мысль:

> Любой человек, кто много и внимательно читает журналы, вынужден признать, что едва ли есть какие-либо препятствия к выходу публикации. Похоже, нет исследования настолько бестолкового, гипотезы настолько заурядной, ссылок на литературу настолько пристрастных или сосредоточенных на самом авторе, дизайна эксперимента настолько извращенного, методов настолько негодных, изложения результатов настолько небрежного, невразумительного или противоречивого, анализа данных настолько необъективного, аргументации настолько несвязной, выводов настолько пустяшных или необоснованных, а грамматики и синтаксиса настолько плачевных, чтобы соответствующая статья не могла быть в итоге опубликована[11].

Он был далеко не первым, кто это заметил. В 1830 году математик и "отец компьютеров" Чарльз Бэббидж написал свои знаменитые "Размышления об упадке науки в Англии и некоторых его причинах", где предложил классификацию недугов науки[12]. Там было "подшучивание" (тех, кто придумывает поддельные открытия, которые затем разоблачает, чтобы доказать свою точку зрения), "фальсификация" (наших пресловутых мошенников от науки, которые не собираются раскрывать свой обман), "обтесывание" и "подделывание" (и то и другое соответствует в нашем современном понимании *p*-хакингу, когда ученые манипулируют данными и наблюдениями, чтобы те выглядели более интересными или точными). Получается, что, хотя современная публикационная система усугубляет проблемы науки, она вряд ли является их первопричиной: эти проблемы с нами уже давно. Вообразите, насколько большего прогресса мы могли бы добиться, насколько больше болезней искоренить или сдержать, насколько больше

узнать о космосе, эволюции, клетке, мозге, человеческом обществе, не говоря уже об огромном количестве ложных надежд и тупиковых путей, которых можно было бы избежать, — сделай мы наконец что-нибудь со всеми этими давнишними неисправностями.

Я задумал написать эту книгу, чтобы поспособствовать процессу изменений, "услужить науке", как выразился в начале своего трактата Бэббидж. Друзья предупреждали его, что критикой научной практики он наживет себе врагов (впрочем, они могли учитывать еще и его довольно сварливый характер)[13]. Когда я рассказывал своим друзьям о будущей книге, их обычно больше беспокоил вопрос доверия к науке: "Не безответственно ли писать что-то подобное? Не сыграешь ли ты на руку тем, кто станет использовать твои аргументы, чтобы оправдать свое неверие в эволюцию, или в безопасность вакцин, или в вызванное человеческой деятельностью глобальное потепление? В конце концов, если фундаментальная наука настолько предвзята, а ее результаты так раздуты, с чего бы обычному человеку верить словам ученых?"

Но рассуждать нужно не так. Во-первых, доверие к науке находится на очень высоком исходном уровне. В рамках исследования *Wellcome Global Monitor* по всему миру собираются данные об отношении людей к науке и ученым. В общемировой выборке 2018 года в среднем 72 % людей сообщили о своем умеренном или высоком уровне доверия к науке, а в Австралии и Новой Зеландии значения достигали аж 92 %[14]. Разумеется, в некоторых регионах уровень доверия был куда ниже: всего лишь 48 % в Центральной Африке, к примеру, и 65 % в Южной Америке. Но среднее значение высоко, и есть свидетельства тому, что в некоторых странах Запада, таких как Великобритания, отношение людей к науке со временем стало *более* доверительным[15]. Во всем мире люди очень уважают науку, и, хотя уровень доверия к ней может немножко снизиться, когда люди услышат о каких-то ее системных проблемах, маловероятно, что наступит резкое катастро-

фическое падение[16]. По моему мнению, ученым нужно усерднее работать, чтобы заслуживать доверие. Мне неизвестны репрезентативные опросы на эту тему, но я уверен, что непрерывная шумиха в средствах массовой информации вокруг раздутых утверждений о научных или медицинских прорывах, равно как и конвейер противоречивых, ненадежных результатов из областей вроде эпидемиологии питания, воздействует на доверие к науке разрушительнее, чем сколько угодно дискуссий о кризисе воспроизводимости.

Впрочем, что важнее, серьезный и глубокий взгляд на науку не подразумевает беспрекословного доверия. Он емко выражен в девизе Королевского общества Великобритании: *nullius in verba*, “не верь ничьим словам”. (Практически та же мысль выражена в русской поговорке, которую любил употреблять Рональд Рейган во время переговоров в период холодной войны: “Доверяй, но проверяй”.) В этом концепция открытой науки, и та же по сути идея лежит в основе мертоновских норм коллективизма и организованного скептицизма: нужно как можно меньше полагаться на бездумное доверие и открывать миру как можно больше проверяемых, верифицируемых, тестируемых доказательств. Есть такая фраза: “На самом деле никакой альтернативной медицины не существует, есть только работающая медицина и неработающая”[17]. Аналогично не существует, в общем-то, никакой “открытой” науки — есть наука, а есть какая-то неведомая, закрытая, недоступная для проверки деятельность, которой занимаются научные сотрудники, и вам остается лишь слепо верить, что они делают все правильно.

Более того, опасно поощрять людей думать о науке как о совокупности фактов, в которых нельзя сомневаться. Такой взгляд на науку не только противоречит норме организованного скептицизма, но и способен привести к очень плохим последствиям. Если вы считаете науку истиной в последней инстанции, в которую вам только и остается что верить, как вы поступите, когда выяснится, что что-то пошло не так?

В конце концов, как мы твердо усвоили из этой книги, в науке довольно часто что-то действительно *идет* не так. Историк науки Алекс Чисар рассматривает проблему изменения климата, где скептики

> создали ложный образ, будто научные публикации — краеугольный камень легитимного консенсуса, только лишь затем, чтобы громко возмущаться, когда окажется, что они не соответствуют этой фантазии. Показательным примером стала реакция на утечку тысяч электронных писем и документов из отдела климатических исследований Университета Восточной Англии в ноябре 2009 года. Выглядело все так, будто электронные письма обличают климатологов, занимающихся в отношении экспертных оценок какими-то скрытными делами и политиканством. За свидетельства того, что научная жизнь за печатным фасадом журналов не является точным отражением ее более благопристойного публичного облика, ухватились многочисленные комментаторы, крича о катастрофе[18].

Так уж получилось, что климатология — подходящий пример. В последние годы эта область подверглась особенно изощренной атаке, когда язык научных преобразований используется как политический ход. В 2019 году министерство сельского хозяйства США объявило своим исследователям, что каждая их работа должна сопровождаться заявлением о ее “предварительности”[19]. Казалось бы, именно это я и рекомендую в книге: рассматривать каждое исследование как осторожный шаг к ответу, а не как ответ сам по себе. Но никто не думает, что эта политика была продиктована невинным желанием улучшить восприятие исследований людьми. Она была объявлена потому, что значительная часть работы, осуществляемой министерством сельского хозяйства, связана с изменением климата, а ее результаты часто неудобны для чиновников, которые выступают за ископаемое топливо, как, например, Дональд Трамп.

Новые правила вызвали бурную реакцию, и через месяц предписание использовать ярлык предварительности было отменено. Впрочем, поразительнее всего было то, как перестарались некоторые ученые, отражая политическую атаку. Газета *Washington Post* процитировала редактора издания *Journal of Environmental Quality*, заявившего, что опубликованная работа — это “конечный продукт вашего исследования… Она уже завершена. В ней нет ничего предварительного”[20]. Это утверждение невежественное, основанное на каком-то идеализированном, приукрашенном представлении о науке, которое мы встречали на страницах этой книги и развенчивали. Даже если политики используют обеспокоенность проблемами с воспроизводимостью как лицемерный предлог для своего скептического отношения к изменению климата, это не оправдывает ученых, завышающих степень доверия, которое мы должны испытывать к нашим результатам. Как отмечают редакторы *Retraction Watch* Айвен Орански и Адам Маркус, “ученым и политикам необходимо руководствоваться долгосрочной перспективой, даже если политики, живущие на ископаемое топливо, хватаются за любую возможность, чтобы поставить под сомнение глобальное потепление. <… > В настоящее время для улучшения воспроизводимости предпринимаются значительные усилия, и… нельзя допустить, чтобы угроза злонамеренных попыток эксплуатировать реформу науки эти усилия подорвала”[21].

Политики уже давно подавляют неудобную для них науку. Самый безумный исторический пример — это, пожалуй, принуждение следовать гротескной псевдонауке Трофима Лысенко, чьи идеи, отрицающие генетику и возведенные в культ, стали одной из причин страшного голода, унесшего миллионы жизней в СССР при Сталине и в Китае при Мао[22]. Однако правительству не обязательно быть тоталитарной диктатурой, чтобы в нем процветала антинаука. Политики в демократических странах регулярно угождают своим избирателям, отрицая или искажая научные данные, будь то американские

политики, поддерживающие преподавание креационизма, итальянские популисты, выступающие против вакцин, правительство Южной Африки, катастрофически отрицающее связь между ВИЧ и СПИДом, или премьер-министр Индии Нарендра Моди, выдающий нелепые теории о том, что технология использования стволовых клеток якобы была доступна еще в Древней Индии[23]. Даже сравнительно либеральное правительство Шотландии в 2015 году наложило запрет на коммерческое выращивание генетически модифицированных культур, а ведь это решение затруднит исследования и вынудит шотландских фермеров обходиться без технологических достижений, например, в области устойчивости к вредителям. Один политический комментатор высмеял такую политику, призванную защитить "целостность" "чистого и зеленого... бренда" Шотландии (что бы это ни значило), и окрестил ее "дешевым популизмом", а в открытом письме, подписанном двадцатью восемью научными обществами, она названа "чрезвычайно тревожной"[24]. Все это говорит нам о том, что, независимо от обсуждавшегося нами кризиса воспроизводимости и связанных с ним провалов, политики все равно будут растаптывать науку, когда это может, по их мнению, добавить им избирательских голосов.

Опасение, что аргументы, приведенные в этой книге, могут быть неправомерно использованы для избирательных лицемерных нападок на исследования, не должно останавливать нас от публичного обсуждения кризиса воспроизводимости и связанных с ним проблем. Мы не должны заставлять науку втягивать живот всякий раз, когда за ней наблюдают представители общественности или политики. На самом деле откровенное признание слабых сторон науки — это лучший способ упредить нападки критиков и в целом честно оценивать, как в действительности работает полный неопределенностей научный процесс.

Аргумент о том, что, вынося сор из избы, мы якобы снижаем доверие к науке, кажется еще более несостоятельным,

когда мы смотрим на огромное количество бесполезных, вводящих в заблуждение и совершенно неблагонадежных исследований, которые выпускаем в мир. Каждый раз, когда мы позволяем опубликовать ошибочное или явно предвзятое исследование, каждый раз, когда мы пишем очередной дутый пресс-релиз, не подкрепляемый данными, каждый раз, когда ученый выдает популярную книгу, полную внушающих оптимизм, но бездоказательных советов, мы вкладываем в руки критиков науки дополнительное оружие. Я бы предложил исцелить науку — и доверие приложится.

Вопреки порочным стимулам, вопреки публикационной системе, вопреки академическому сообществу и ученым, наука содержит в себе инструменты для самоизлечения. С помощью самой же науки мы можем выяснить, где наши исследования пошли не так, и придумать, как все исправить. Проблема не в идеалах научного процесса — проблема в том, что мы предаем эти идеалы при проведении исследований на практике. Если мы сумеем хотя бы начать приводить практику в соответствие с ценностями, то сможем вернуть все утраченное доверие — и с чистой совестью отступить в изумлении перед всеми чудесными открытиями.

Эмиль Золя определил художественное произведение как “кусок действительности, увиденный сквозь темперамент”*[25]. Как мы снова и снова убеждались на протяжении всей книги, это определение с равным успехом можно применить и к науке — или, во всяком случае, к тому способу, каким она сейчас делается. Куски действительности, с которыми наука имеет дело, видятся сквозь слишком человеческие темпераменты с их сопутствующими предрассудками, высокомерием, небрежностью и нечестностью. Даже не считая, что наука — всего лишь одна из многих равноправных “истин”,

* Перевод В. Шора.

нельзя не согласиться с тем, что это определенно человеческая деятельность, а значит, она несет на себе отпечаток человеческих недостатков.

Столь грандиозная задача, как преображение науки, и не может быть простой. На этом пути громоздятся пробы, ошибки и, соответственно, эксперименты. Речь идет не только об отбрасывании какой-либо ошибочной теории вроде геоцентризма, или флогистона, или алхимии, или любого другого экспоната из паноптикума неверных идей, коими пестрит история науки. Речь идет о коренном изменении нашего способа проводить исследования и всей научной культуры — о попытке преодолеть изъяны и предубеждения, которые прокрались в науку, по большей части незаметно. Мир по праву гордится тем, куда привела нас наука. Ради сохранения этого чувства гордости мы должны ей нечто куда большее, чем продукт наших несовершенных человеческих темпераментов.

Мы должны ей правду.

Приложение
Как читать научную статью

Помните наказ Даниэля Канемана из второй главы о том, что “вам придется” поверить в поведенческие эффекты “прайминга” — которые позднее при попытках их воспроизвести показали свою несостоятельность? Основная идея этой книги такова: сталкиваясь с научным результатом, вы *обладаете* свободой выбора — вы вправе решить воздерживаться от высказывания своего мнения до тех пор, пока тщательным образом не проверите, как и откуда этот результат взялся.

Но как? Бесспорно, чтобы полностью понять сильные и слабые стороны какого-либо исследования, обычно требуются годы подготовки в соответствующей научной области. И тем не менее если вы решите провести небольшой поиск в интернете, то зачастую сумеете получить необходимое представление о достоинствах и недостатках работы. А если пожелаете проверить саму статью, есть набор красных флагов, которые могут проглядывать даже сквозь непроницаемый технический жаргон.

Разумеется, проверка интересующей вас статьи предполагает, что вы можете ее скачать, правда, полный открытый доступ к научной литературе пока не стал реальностью, так что публикация, вероятно, окажется платной. С этим можно кое-что попытаться сделать, если вы, конечно, не хотите просто выложить деньги за доступ к статье — с чем совершенно

точно не следует торопиться, пока вы не испробовали все остальные варианты, которые я сейчас опишу (а когда они провалились, не расхотели вообще читать эту работу). Во-первых, имеет смысл проверить, не выложена ли на личных или рабочих сайтах авторов бесплатная версия для скачивания, иногда в несверстанном виде. Обычный поиск статьи в *Google Scholar*, когда вы нажимаете под каждым элементом из списка выдачи на “версии” документа, часто может привести вас к этим бесплатным рукописям статей. Во-вторых, можно проверить, не был ли выложен препринт — они всегда выкладываются в открытом доступе, и хотя иногда препринт немного отличается от финальной, публикуемой в журнале версии, которая прошла этап рецензирования, чаще всего препринт и статья очень похожи. (Как мы видели, бывает, что именно препринт попадает в основные средства массовой информации, откуда вы, возможно, и узнали об исследовании.) В-третьих, можно написать по электронной почте авторам: хоть мы и обсуждали, что ответы приходят печально редко, когда люди, намеревающиеся воспроизвести исследование, просят у авторов их данные, однако просьба поделиться статьей вряд ли вызовет проблемы (на самом деле многим ученым будет скорее приятно: кого-то настолько заинтересовало их исследование, что он даже готов прочитать статью). В-четвертых, поговаривают, что еще есть разные пиратские способы, позволяющие добыть статью в интернете нелегально, но об этом я ничего не знаю[1].

Допустим, вы заполучили полную версию статьи. При ее прочтении вы можете применить почти все знания, почерпнутые из этой книги, используя приведенный ниже список из десяти главных вопросов.

1. *Все ли прозрачно?* Во-первых, основное: авторы из вроде бы заслуживающих доверия университетов, компаний и лабораторий? Журнал, где опубликована работа, выглядит профессиональным? Если сайт издания ужасен, словно из ка-

ких-нибудь 90-х, это, вероятно, один из удручающе распространенных “хищнических” журналов, которые встретились нам в седьмой главе. Ничему из опубликованного в подобных местах доверять не стоит, поскольку там даже и не *пытаются* рецензировать статьи[2].

2. *Насколько все открыто?* Иными словами, насколько исследование отвечает идеям открытой науки, описанным в восьмой главе? Было ли оно предварительно зарегистрировано? Утвердительный ответ никоим образом не гарантирует, что результаты верны, так же как ответ отрицательный не означает, что они ошибочны. Но если вы сумели отыскать онлайн-регистрацию исследования, то можете быть по крайней мере немного увереннее, что результаты обусловлены не только лишь *p*-хакингом[3]. Просматривая документы предварительной регистрации, вы также можете определить, отличается ли анализ результатов в итоговой статье от ранее заявленного: то есть не прибегли ли ученые к переключению на другие исходы? А еще: доступны ли данные и другие материалы онлайн? Как мы обсуждали, не каждый набор данных возможно сделать общедоступным, вдруг он, например, содержит информацию, которая позволит идентифицировать конкретных участников исследования. Впрочем, такие случаи редки. Если есть ссылка на полный набор данных, которую легко найти, это убедительное свидетельство в пользу того, что ученые с читателем откровенны[4].
3. *Правильно ли исследование спланировано?* В пятой главе мы говорили, что в настораживающе большой доле статей, посвященных работе на животных, рандомизация и ослепление даже не упоминаются. А это существенные аспекты в дизайне эксперимента, так что, если в статье они никак не обсуждаются — как минимум в работах, посвященных клиническим испытаниям, где эти аспекты критически важны, — ваша подозрительность должна усилиться. Кроме того, для многих исследований необходима подходящая контрольная группа. Когда ваш взгляд цепляется в статье за сенсационное утвер-

ждение, всегда надо задаваться вопросом “по сравнению с чем?”. Если ответ — “по сравнению с контрольной группой, отличавшейся от лечебной по важным параметрам еще до начала эксперимента”, то перед вами плохо спланированное исследование.

4. *Насколько велика выборка?* Размер выборки имеет значение, главным образом из-за статистической мощности. Это правда, что статистическую мощность можно увеличить и другими способами, поэтому размер выборки — не единственный фактор. Для некоторых видов исследований, где, например, ожидаются большие эффекты или участники тестируются снова и снова, скромной выборки совершенно достаточно. И даже огромные выборки бывают безнадежно смещенными, если они неслучайны или нерепрезентативны. Но для исследований в таких областях, как нейронаука, экология и психология, это распространенная ошибка — искать в крошечных выборках в принципе слабые эффекты; такая стратегия оказывается хуже, чем просто бесполезной. Еще нужно следить за тем, сколько испытуемых оказались исключены из финальной выборки. Некоторые исключения абсолютно в порядке вещей и обычно даже неизбежны. Например, люди, участвующие в исследовании, редко все поголовно следуют указаниям, как должны были бы. Тем не менее, если исключений слишком уж много, скажем больше половины выборки, вы вправе засомневаться, можно ли обобщать результаты на изучавшуюся популяцию, или авторы просто отобрали тех участников, которые демонстрировали желаемый эффект, а остальных выкинули.

5. *Насколько велик эффект?* Первое, что надо проверить: статистически значим ли заявленный в статье эффект и каков уровень значимости. Высчитано ли много p-значений, которые чуть ниже порога в 0,05? Используют ли авторы туманные выражения вроде “тенденция в сторону значимости”, дабы сгладить тот факт, что их результаты недостаточно хороши? Впрочем, это лишь начало — еще вы должны задаться

вопросом, насколько велик обнаруженный в исследовании эффект. Как он соотносится с другими исследованиями или с другими релевантными эффектами? К примеру, если в работе анализируется новое медицинское вмешательство или образовательная инициатива, как результаты соотносятся с другими, уже устоявшимися медицинскими методами и образовательными практиками? Не интерпретируется ли средствами массовой информации или самими учеными какой-то небольшой эффект так, словно он единственное, что имеет значение? Поскольку мы знаем, что на ту же тему может существовать несколько исследований с отрицательными результатами, запрятанных в "картотечный ящик", полезно в уме немного уменьшить величину эффекта. Проблема в том, что неправдоподобно большие эффекты — которые, попросту говоря, слишком хороши, чтобы быть правдой, — также должны вызывать подозрение, что с исследованием не все в порядке. То же касается и *p*-значений: при виде исследования, сообщающего исключительно или почти исключительно о статистически значимых результатах, стоит удивленно выгибать бровь. Ибо, как мы видели, исследования никогда не обладают идеальной статистической мощностью, а часто она на самом деле очень низка. Поэтому, даже если заявленные эффекты правдивы, все равно логично ожидать, что некоторые *p*-значения не перевалят через порог 0,05. Ровный строй статистически значимых результатов в исследовании, где вычислялось много *p*-значений, намекает на *p*-хакинг (или на что похуже).

6. *Адекватны ли выводы?* Как мы видели, ученые регулярно впадают в "причинно-следственный" тон, даже если провели всего лишь корреляционное исследование. Если ученые на основании результатов наблюдательного исследования говорят о том, как переменная X влияет или оказывает воздействие на переменную Y, они выходят за пределы своих данных. В наблюдательном исследовании нет никакого рандомизированного вмешательства, поэтому выводов о при-

чинно-следственных связях обычно сделать нельзя. Аналогично, если эксперимент выполнен на мышах или крысах либо на компьютерной модели, заключение, что такой опыт обязательно дает нам некую информацию о том, “как все устроено у людей”, попросту неверно. То же касается исследований, проведенных лишь на небольшой выборке людей, но представленных так, будто результаты сообщают нам что-то о человечестве в целом.

7. *Есть ли предвзятости?* Нет ли у исследования явной политической или социальной подоплеки? И не сообщают ли о ней ученые кажущимся не слишком уж беспристрастным образом? Мы видели, что преувеличения и подача под нужным углом часто ужасающи, даже в рецензируемых статьях. Не профинансировано ли исследование — полностью или частично — каким-либо коллективом или компанией, для которых предпочтительным был бы один конкретный исход? Некоторое представление об этом можно получить, заглянув в разделы “Источники финансирования” и “Конфликт интересов”, обязательные почти для всех журналов (но имейте в виду, что сейчас там не требуется упоминать о таких вещах, как контракты с издательствами и лекционные туры, а ведь они могут быть напрямую связаны с результатами статьи; вероятно, вам захочется проверить подобный второстепенный вид деятельности авторов на их сайтах). Если авторы подобающе осторожны при обсуждении своих открытий и не отыскиваются материалы средств массовой информации, где эти ученые говорили бы журналистам, как их результаты отчетливо согласуются с некой политической позицией или конкретным политическим курсом, это хороший знак, свидетельствующий о том, что они сдерживают свои предубеждения. Между прочим, еще важнее отслеживать предвзятости, когда исследование согласуется с вашими *личными* идеологическими предубеждениями. Стоит спросить себя: я подвергаю исследование несоразмерно жесткой проверке, поскольку не согласен с его выводами, или же не глядя при-

нимаю слабую работу, потому что в ней подкрепляются мои предрассудки?

8. *Насколько все вообще правдоподобно?* Если речь идет об исследовании с участием людей, полезно представить, будто вы сами были его участником[5]. Например, в случае работы по эпидемиологии питания подумайте, насколько точно при заполнении опросника по частоте потребления различных пищевых продуктов вы сумели бы вспомнить свои привычки, касающиеся перекусов, за последние десять лет или хотя бы за несколько последних недель. Наверняка ответ — “не очень уж точно”. Не выматывались бы вы к концу всех тестирований в поведенческом эксперименте и учитывалось ли это исследователями? Точно ли обстановка, в какой проводится исследование (скажем, лаборатория в университете), имитирует условия, которые, собственно, и интересуют ученых (например, как при собеседовании на высокую должность)? Иными словами, действительно ли исследование отвечает поставленной задаче? Когда вы ставите себя на место участника, это помогает определиться с подобными базовыми вопросами, касающимися правдоподобности работы.
9. *Воспроизводилось ли исследование?* Нам нужно прекратить всецело доверять отдельным работам. Надежнее, когда ученые повторно получили собственные результаты, а еще лучше, если их воспроизвели и другие исследователи из совершенно независимых лабораторий. Первое, что надо сделать, — поискать, не опубликованы ли какие-то исследования-повторения[6]. Еще может найтись обзор или метаанализ, посвященный основным результатам статьи или же сходным результатам, который покажет, не является ли интересующее вас исследование просто некой аномалией, а также вписываются ли его результаты в более общую теорию (помните, что заинтересовавшее вас исследование само может быть неудавшейся попыткой воспроизвести более ранний результат). Разумеется, обзоры и метаанализы и сами бывают искажены

из-за публикационного смещения и плохого качества исходных работ; если вам удалось найти метаанализ исследований, каждое из которых было заранее зарегистрировано, вы сорвали джекпот — правда, сам я, кажется, никогда не сталкивался с таким уникальным случаем (хотя в будущем это может измениться, поскольку предварительная регистрация становится все популярнее). Естественно, нельзя рассчитывать на то, что существует работа, воспроизводящая интересующее вас исследование, если оно совсем свежее и новаторское, однако можно воздержаться от суждений о его достоверности, пока такая работа не появится.

10. *Что думают об этом исследовании другие ученые?* В лучших новостных материалах о научных работах обычно цитируются слова какого-нибудь независимого ученого, выражающего свое мнение, так что имеет смысл посмотреть, нет ли уже где-то отзывов. Есть и организации, занимающиеся этим систематически: например, британская благотворительная организация “Научный пресс-центр” всякий раз, как выходит пресс-релиз о новой статье, запрашивает комментарии и отзывы у многих независимых экспертов и публикует на своем сайте[7]. Это хороший пример того, как рецензирование может происходить даже после официального выхода статьи. Еще стоит поискать информацию на сетевых ресурсах вроде *Pubpeer*, анонимного сайта с комментариями к научным статьям, где впервые вскрылось, что изображения из публикации о стволовых клетках Харуко Обокаты подделаны, и где разоблачались хитрости еще многих других мошенников[8]. Поискать какие-нибудь блоги или любые сайты, где обсуждается статья, в *Google* или даже пошерстить в твиттере тоже бывает полезно, только нужно отдавать себе отчет, что таким образом можно обнаружить как грамотное, так *и* бестолковое, как серьезное, так *и* шутливое, как непредвзятое, так *и* предвзятое обсуждение исследования[9]. Если работа вышла не вчера, вы можете использовать функцию “Цитируется” в *Google Scholar*, чтобы посмотреть, есть ли на эту ста-

тью ссылки, и проверить, как ее чаще цитируют — в положительном или же отрицательном ключе[10].

Ни один из этих довольно общих приемов не идеален, и не все они применимы к каждому виду исследований. Очевидно, что всегда лучше обладать какими-то предварительными знаниями и опытом в определенной научной области, чтобы полнее оценить сильные и слабые стороны конкретного исследования. Впрочем, все лучше, чем просто принимать чьи-то утверждения за чистую монету.

Еще важно помнить, чему мы научились благодаря саге о Стивене Джее Гулде, Сэмюэле Мортоне и нескончаемых дебатах о размерах черепов: даже если вы читаете кажущуюся разгромной критику некоего исследования, сам критический разбор может оказаться ошибочным, как и критика критического обзора. Это относится и ко всему, о чем я написал в этой книге.

Фундаментальный урок таков: нужно *скромнее оценивать то, что мы знаем и чего не знаем*. На первый взгляд может показаться, что это противоречит идее научного исследования, которая определенно в том и состоит, чтобы открывать новые факты о мире и всегда наращивать знание. Но поразмыслите хорошенько — и вы обнаружите, что сама суть науки в этом.

Благодарности

Все началось с Уилла Фрэнсиса, моего литературного агента, который и предложил мне написать книгу на эту тему. Уилл, Пи Джей Марк и вся команда *Janklow & Nesbit* невероятно помогали мне на всех этапах создания текста.

Работая с моими редакторами, Уиллом Хэммондом и Григорием Товбисом, я частенько вспоминал старое клише о том, что редактирование — это процесс превращения уродливой глыбы мрамора в красивую скульптуру. Поверьте, вам бы не захотелось читать эту книгу (если ее можно было бы так назвать), не внеси они свою вдумчивую и тщательную правку. Выражаю благодарность также сотрудникам Уилла и Григория в издательствах *The Bodley Head* и *Metropolitan Books* (особенно Элисон Дэйвис и Саре Фиттс) за все их усилия ради того, чтобы книга увидела свет. Еще я благодарен Мэриголд Этки за тщательную и благодушную правку, а Генри Кауфману — за обнадеживающую юридическую консультацию.

Несколько моих друзей читали и комментировали различные черновые варианты, и я очень им благодарен. Это Ник Браун, Ива Чукич, Джереми Драйвер, Стейси Шоу, Крис Сноудон и Кэти Янг. Отдельное спасибо хочу сказать двум читателям: Салони Даттани, которая моментально прочитывала первые черновики каждой главы, как только я их писал, и тут же сообщала, как они выглядят и что получилось, а что нет, и Анне Шеель, которая вышла далеко за рамки своих обязанностей

и спасла меня от многочисленных статистических (и других) ловушек благодаря своему знанию статистики, необъятным познаниям в области открытой науки и исключительной проницательности по части формулировок и разных нюансов.

Я также в долгу перед всеми, кто подсказывал мне новые истории или ссылки, вел интересные беседы или спорил со мной о науке и ее проблемах либо просто оказывал столь необходимую поддержку в процессе написания книги. Это все члены группы *Best Picture* (Бобби Блюбелл, Кенни Фаркухарсон, Юэн Макколм и Иэн Рэнкин), все не упомянутые еще члены группы *Fat Cops* (Крис Эйр, Крис Дирин, Эл Мюррей и Нил Мюррей), Мораг и Найджел Аткинсоны, Майк Бёрд, Эван Бёрни, Робин Биссон, Сэм Боуман (который, как ни досадно это признавать, подал мне идею для названия книги), Крис Чабрис, Том Чиверс, Саймон Кокс, Гейл Дэйвис, Иэн Дири, Рори Эллвуд, Аласдер Фергюсон, Патрик Форшер, Анна Фёртджес, Роджер Джинер-Соролла, Ниалл Гуч, Саския Хагенаарс, Сара Хэйдер, Льюис Хэлси, Пейдж Харден, Кирсти Джонсон, Майк Джонс, Мустафа Латиф-Арамеш, Риккардо Мариони, Дэмиен Моррис, Ник Партингтон, Роберт Пломин, Дженнифер Рафф, Джо Роулинг, Адам Резерфорд, Илвин Скалли, Эдриан Смит, Бен Саутвуд, Майкл Стори (и собака Ласка), Эллиот Такер-Дроб, Симин Визайр, Рэйчел Вагнер, Эд Уэст, Сэм Вествуд, Таль Яркони и сотрудники Центра научных СМИ. Благодаря моим коллегам из Центра социальной, генетической и возрастной психиатрии в Королевском колледже Лондона для меня было истинным удовольствием каждый день работать, а потом возвращаться домой и снова садиться за книгу. Их слишком много, чтобы всех называть по отдельности, но особого упоминания заслуживают мои замечательные начальники Фрэнки Хаппе и Кэтрин Льюис, которые очень меня поддерживали.

Разумеется, перечисленные выше (или ниже) люди не обязательно согласны со всеми моими мыслями, изложенными в книге, и никто из них не несет ответственности за лю-

бые ошибки, вызванные моими собственными предубеждениями или небрежностью. Кстати, я призываю всех, кто обнаружит какую-либо ошибку в тексте, связаться со мной через сайт *sciencefictions.org* и указать точную информацию. Все исправления я публикую на этом же сайте.

Многие, слишком многие люди рассказывали мне (часто шокирующие) истории о мошенничестве, предвзятости, недобросовестности и хайпе в науке, которые они раскрыли, а иногда и прожили сами, будучи студентами или научными сотрудниками в лабораториях, где все было очень неправильно. Жаль, что я сумел вместить в текст лишь малую часть этих историй, и могу только надеяться, что моя книга вдохновит тех, кто знает примеры недобросовестного проведения исследований или даже честных, но ошибочных работ, сделать все возможное, чтобы поведать об этом миру.

Писать книгу — дело непростое, и я могу лишь извиниться перед всеми коллегами, которые удивлялись, какого черта я так долго отвечал им по тому или иному проекту, и перед друзьями, во время самых напряженных этапов моей работы наверняка думавших, что я как в воду канул. Мои родители, как всегда, были ко мне невероятно добры и на протяжении всего процесса подбадривали — я не в силах выразить, как им благодарен. Но больше всего написание книги ударило по Кэтрин Аткинсон. И тем не менее, когда я в очередной раз повторял “Мне нужно вернуться к книге… ”, она проявляла лишь терпение, не показывая ни капли разочарования. Я определенно не заслужил столько великодушия. Именно ей посвящается книга.

Стюарт Ричи
Март 2020 года

Послесловие

Я закончил писать книгу “Наукообразная чушь” в начале 2020 года, поэтому в тексте есть лишь одно краткое упоминание о том, что станет крупнейшей научной историей наших жизней, — о пандемии COVID-19. Со времен холодной войны и космической гонки мы не слышали о науке на ежедневной основе больше и не возлагали на нее столько надежд. К счастью для “Наукообразной чуши” — но к большому несчастью для всего мира, — пандемийная наука страдала от тех же самых недостатков, что я описал в своей книге.

Не думаю, что с моей стороны это просто ошибка подтверждения: примерно та же история, которую я описал в эпилоге, — о потоке удивительных открытий, загрязненном обманчивыми, предвзятыми и небрежными исследованиями, — разыгралась во время пандемии, только в ускоренном темпе. С одной стороны, в чем-то мы добились невероятного прогресса, например в создании мРНК-вакцин, которые спасут бесчисленное количество жизней и помогут в борьбе со многими другими заболеваниями помимо COVID-19[1]. С другой стороны, мы видели пачки низкокачественных научных статей, которые порождали смятение среди медицинских работников, растрачивали деньги и время впустую и оказывали поддержку тем, кто занимал “противоположную” позицию в отношении опасности вируса.

О научных спорах, связанных с пандемией, могут быть и будут написаны целые книги: смертоносность вируса и механизм его передачи, длительность (и сама идея) локдаунов, значение “коллективного иммунитета”, закрытие школ, использование масок для лица, особенности различных вакцин — все это вызвало жесткие разногласия среди ученых[2]. Но сейчас давайте кратко проследим историю одного научного спора, в котором ни одна из сторон не выглядит достойно и который иллюстрирует ряд уже знакомых нам проблем. Речь идет о страстях вокруг гидроксихлорохина[3].

Гидроксихлорохин — это препарат, который совершенно точно работает, а именно при малярии. Вопрос, возникший в самом начале пандемии, заключался в том, может ли это лекарство помогать и при COVID-19. В марте 2020 года исследование микробиолога и врача из Университета Экс-Марсель Дидье Рауля и его коллег ясно показало “стопроцентную частоту излечения” больных коронавирусом, принимавших этот препарат вместе с антибиотиком азитромицином[4]. Рауля в его лаборатории посетил президент Макрон, а по другую сторону Атлантики президент Трамп объявил гидроксихлорохин “переломным фактором”[5].

Разумеется, то был хайп. На самом деле исследование не только не соответствовало важным критериям, которые мы обсуждали в пятой главе (не было ослепления, должной рандомизации, а размер выборки был очень мал), но и имело ряд уникальных и серьезных недостатков. Пожалуй, самый важный из них заключался вот в чем: хотя трое пациентов с COVID-19, получавших гидроксихлорохин, оказались в итоге в реанимации, а один умер, ученые об этих исходах не сообщили, поскольку сосредоточились исключительно на мазках, по которым определялось количество вируса в носу у каждого пациента. По словам журналистки Джулии Кэрри Вон, “поскольку [четыре] пациента находились в отделении интенсивной терапии или умерли, образцы у них взять не могли — и в окончательный анализ они включены

не были"[6]. И вуаля: ученые смогли заявить, что, *согласно проанализированным ими данным*, препарат прекрасно работает[7].

Несмотря на такой мухлеж, исследование все равно вышло и вызвало ажиотаж в области изучения гидроксихлорохина: в какой-то момент по всему миру проводились сотни отдельных испытаний — совершенно разного качества, — в ходе которых ученые пытались понять, действительно ли перед ними чудодейственное средство[8]. Сейчас, кстати, мы можем сказать, что, согласно результатам наиболее крупных и надежных исследований, гидроксихлорохин не помогает в лечении COVID-19[9].

Одно из исследований, запущенных под влиянием работы Рауля, было проведено учеными из Бригамской женской больницы, входящей в состав Гарвардской медицинской школы. Они провели работу в сотрудничестве с медицинской аналитической компанией *Surgisphere*, имевшей доступ к набору данных о многих тысячах пациентов с COVID-19 по всему миру. В отношении гидроксихлорохина данные были однозначны и не сводились лишь к тому, что "гидроксихлорохин не работает", — на самом деле пациенты, которых лечили этим препаратом, *умирали чаще*. Исследование было корреляционным, а не рандомизированным, но результаты тем не менее выглядели важными — настолько, чтобы в мае 2020 года выйти в таком престижном журнале, как *The Lancet*[10].

То были ценные новости из вроде бы надежного источника. В ответ Всемирная организация здравоохранения приостановила свои испытания гидроксихлорохина, а Франция изменила рекомендации для врачей[11]. Однако, как и в случае с другими резонансными исследованиями, которые мы обсуждали в этой книге, привлечение внимания мировой общественности породило несколько серьезных вопросов о гарвардской работе. Каким образом *Surgisphere* — крошечная компания с горсткой сотрудников — сумела так быстро получить столько данных из такого количества больниц?[12] Почему ни одна из этих больниц не была названа или упомянута в ста-

тье? И вообще: не слишком ли велик эффект, реалистичен ли? Расследователи в мире данных начали разбираться с числами и обнаружили некоторые странные несоответствия[13].

Вот тут-то и вскрылась недобросовестность. Когда гарвардским ученым задали эти вопросы, они признались, что сами не изучали данные. Они просто доверили *Surgisphere* провести анализ и, ознакомившись с результатами, радостно поставили в статье свои имена перед отправкой ее в журнал. А теперь *Surgisphere* отказывалась предоставить им исходные данные, заявляя, что это нарушило бы соглашения о конфиденциальности. Оставался только один выход: *The Lancet* пришлось отозвать статью — всего через две недели после публикации[14].

Свидетельствовали ли сомнительные данные компании *Surgisphere* о научном мошенничестве? В отличие от многих других случаев, рассмотренных в третьей главе, официального расследования еще не проводилось (все осложняется тем, что это частная компания, а не университет). Мы попросту не знаем, в чем там дело — имела ли место крайняя некомпетентность или ошибка, а то и нечто более зловещее. Как бы то ни было, эта история служит классическим примером того, как спешка с публикацией — не *только* ради резюме, но и во имя полезного открытия во время пандемии — может выставить дураками даже лучших исследователей и самые престижные журналы.

Похоже, впредь журнал *The Lancet* будет менее доверчив. После фиаско с *Surgisphere* его редакторы объявили о планах изменить правила рецензирования статей. Теперь, принимая на рассмотрение работы, в которых используются “большие данные”, журнал будет следить за тем, чтобы один из рецензентов обладал опытом проведения необходимых анализов[15]. Казалось бы, в столь уважаемом издании давно должны были придерживаться подобной практики, но лучше поздно, чем никогда.

В то время как я пишу эти строки, Дидье Рауль предстал перед дисциплинарной комиссией после серии жалоб

от своих французских коллег-клиницистов на его “неэтичное” продвижение гидроксихлорохина[16]. Если ему вынесут взыскание, это будет ощутимым ударом для одного из самых цитируемых ученых в мире, имеющего огромный индекс Хирша, равный 187 (этот показатель обсуждался в седьмой главе), и бесконечно длинный список публикаций (например, он соавтор трети из *всех* семисот двадцати восьми статей, опубликованных за всю историю одного журнала по микробиологии[17]). Если вы задаетесь вопросом, как же ученый с таким завидным, впечатляющим резюме может заниматься низкокачественными, раздутыми исследованиями, предположу, что вы невнимательно читали последние несколько глав этой книги.

В условиях пандемии последнее, что нам нужно, — это плохая наука[18]. Но случилось так, что известные, авторитетные исследователи из уважаемых институтов опубликовали в ведущих мировых журналах работы, которые увели нас в абсолютно неверном направлении. Плохо продуманные никчемные исследования, сомнительные данные, громкие заявления, не подкрепленные доказательствами, доверие среди ученых, переходящее в легковерие, и многое другое — сложно было бы придумать более полное повторение всех тем, затронутых в этой книге. Хоть мне и неприятно об этом сообщать.

Научный процесс, вместо того чтобы работать как положено, неустанно отлавливая ошибки и продвигаясь к истине, для гидроксихлорохина оставил распахнутой дверь, через которую ошибки хлынули в литературу, а оттуда — на столы высокопоставленных политиков. А тем временем люди, на благо которых должна была работать наука, — пациенты с COVID-19, умирающие тысячами, и их врачи, отчаянно нуждающиеся в достоверной информации о методах лечения, — оказались брошены в беде.

Хотя неудачи с исследованиями, посвященными COVID-19, наиболее ярко проиллюстрировали проблемы,

о которых я рассказал в этой книге, в прошедшем году не было недостатка и в плохой науке, не имеющей отношения к пандемии. Если бы я сейчас принимался за эту книгу, то мог бы написать о том, как инженер-исследователь Али Назари сместил Дидерика Стапела с пятого места в "списке лидеров" *Retraction Watch*, когда шестьдесят одна его статья (против пятидесяти восьми у Стапела) была отозвана за отъявленный плагиат и поддельные рецензии*[19]. Или о том, что такая захватывающая и футуристическая научная область, как квантовые вычисления, переживает свой собственный кризис воспроизводимости[20]. Или что исследователи продолжают строить модели, описывающие, как *p*-хакинг и публикационное смещение искажают научную литературу[21]. Или даже рассказал бы сюрреалистическую историю, как один кардиолог опубликовал в научном журнале *Early Human Development* девятнадцать статей, посвященных телесериалу "Звездный путь"[22].

И хотя эти примеры не столь ужасны, как сага о гидроксихлорохине, все они демонстрируют, что в науке практически везде, куда ни глянь, можно обнаружить те же проблемы с мошенничеством, предвзятостью, недобросовестностью и хайпом, а также порочные стимулы, которые их порождают.

Поиск эффективных способов для исправления плачевной ситуации с нашей научной системой должен стать главным приоритетом в любой области исследований. Раз все мы находимся в одной лодке, то, вероятно, сможем учиться друг у друга — и это вселяет надежду.

Стюарт Ричи
Май 2021 года

* По состоянию на июль 2023 года Али Назари со шлейфом из ста отозванных статей воцарился на пятом месте списка, а Дидерик Стапел с пятьюдесятью восемью сместился на седьмое.

Источники иллюстраций

Автор благодарен за разрешение воспроизвести изображения из следующих источников:

с. 55 — Сидни Харрис, *ScienceCartoonsPlus.com;*
с. 249 — свободные материалы *Getty*;
с. 253 — издательство Кембриджского университета;
с. 297 — Джоэл Петт.

Рисунки на с. 84 и 269 воспроизводятся по лицензиям *CC-BY* и *CC-0 Creative Commons* соответственно.

Примечания

Предисловие

1 Morris C. et al. *Paedogeddon!* // *Brass Eye.* Series 2. Episode 1. 26 July 2001.

2 Bacon F. *Novum Organum*. New York: P. F. Collier & Son, 1620/1902. [Бэкон Ф. *Новый Органон* // Соч. в 2-х т. Т. 2. М.: Мысль, 1972.]

3 Bem D. J. *Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect.* Journal of Personality and Social Psychology. 100, no. 3 (2011): 407–25.

4 Студенты также обладали обратной способностью: когда за одним из занавесов скрывалась жестокая картинка, они психически дистанцировались от нее, выбирая соответствующее изображение лишь в 48,3 % случаев — отличие от результата, который должен был бы получиться при чистом угадывании, опять-таки оказалось статистически значимым.

5 Aldhous P. *Journal Rejects Studies Contradicting Precognition*. New Scientist. 5 May 2011.

6 The Colbert Report. *Time Travelling Porn — Daryl Bem*. 2011.

7 После еще нескольких отказов мы в конце концов опубликовали статью в другом научном журнале: Ritchie S. J. et al. *Failing the Future: Three Unsuccessful Attempts to Replicate Bem's "Retroactive Facilitation of Recall" Effect.* PLOS ONE. 7, no. 3 (2012): e33423. Примечательно, что отклонивший нашу статью журнал опубликовал критические замечания, касающиеся статистического анализа в исследовании Бема (Wagenmakers E.-J. et al. *Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011).* Journal of Personality and Social Psychology. 100, no. 3 (2011): 426–32), а также ответ Бема и его коллег (Bem D. J. et al. *Must psychologists change the way they analyze their data?* Journal of Personality and Social Psychology. 101, no. 4 (2011): 716–9). И все равно редколлегия журнала не собиралась рассма-

тривать возможность публикации статьи о попытке повторить исходное исследование. Дальше мы увидим, что редакторы журнала с тех пор пересмотрели свое отношение к этой важнейшей проблеме.

8 Stapel D. A., Lindenberg S. *Coping with Chaos: How Disordered Contexts Promote Stereotyping and Discrimination*. Science. 332, no. 6026 (2011): 251–3.

9 Ball P. *Chaos Promotes Stereotyping*. Nature. 7 April 2011; Phillips N. *Where There's Rubbish There's Racism*. Sunday Morning Herald. 11 April 2011.

10 Stapel D. A., Lindenberg S. *Coping with Chaos.*

11 Levelt Committee et al. *Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel [English Translation]*. 28 Nov. 2012.

12 Stapel D. A. *Faking Science: A True Story of Academic Fraud*. Tr. Brown N. J. L. Strasbourg, France, 2014, 2016. *http://nick.brown.free.fr/stapel*

13 Там же.

14 Действительно, научный прогресс зависит от нашей способности находить допущенные ранее ошибки и недоработки. Например, в начале XX века физики осознали, что классическая механика Ньютона, долгое время считавшаяся универсальной, не может объяснить поведение очень маленьких и очень быстрых частиц, и тогда разработали квантовую механику. Обсуждение этого вопроса с точки зрения измеримых величин вроде скорости света и постоянной Планка см. здесь: Milton M. J. T., Possolo A. *Trustworthy Data Underpin Reproducible Research*. Nature Physics. 16, no. 2 (2020): 117–9.

15 Я нашел единственный пример — вот эту статью, которая была опубликована спустя долгое время после разоблачения мошенничества Стапела, хотя само исследование проводилось до того: IJzerman H. et al. *Perceptual Effects of Linguistic Category Priming: The Stapel and Semin (2007) Paradigm Revisited in Twelve Experiments*. Acta Psychologica. 157 (2015): 23–9. Больше о затяжных попытках опубликовать это исследование (что кое в чем напоминает мой собственный опыт с работой Бема) написано здесь: Zwaan R. *When Replicating Stapel is not an Exercise in Futility*. Zeitgeist. 18 Jan. 2015.

16 Цитируется по: Engber D. *Daryl Bem Proved ESP Is Real: Which Means Science Is Broken*. Slate. 7 June 2017.

17 Классический пример такой книги: Sagan C. *The Demon-Haunted World: Science as a Candle in the Dark*. New York: Ballantine Books, 1997. [Саган К. *Мир, полный демонов. Наука — как свеча во тьме*. М.: АНФ, 2021.]

18 В этой книге я рассматриваю множество сучков в глазах других ученых, поэтому — если вы простите мне минутку самоанализа — правильно будет проверить наличие любых потенциальных бревен в собственном глазу. За годы, прошедшие с моих попыток воспроизвести результаты Бема, у меня вышло много статей на разные темы, хотя преимущественно все-таки про человеческий интеллект, интересующий меня больше всего. В первую очередь нужно сказать, что я никогда умышленно не подделывал результаты. Однако глупо было бы думать, что я неуязвим для предубеждений. Они часто — или, возможно, обычно — бессознательны, и история некоего исследования легко переписывается так, что начинает казаться, будто исходно вы именно таким его и задумывали. Хорошо, что я опубликовал немало отрицательных результатов, то есть статей, в которых не было найдено подтверждений основной гипотезе. См., например: Ritchie S. J. et al. *Polygenic Predictors of Age-Related Decline in Cognitive Ability.* Molecular Psychiatry. 25 (2020): 2584–98; а также мою самую первую научную публикацию: Ritchie S. J. et al. *Irlen Colored Overlays Do Not Alleviate Reading Difficulties*. Pediatrics. 128, no. 4 (2011): e932–8. Опять-таки мне легко можно возразить, что в этой моей дебютной статье с отрицательными результатами изучалась слишком маленькая выборка и потому реальные эффекты могли быть упущены (см. обсуждение статистической мощности в пятой главе). Некоторые мои статьи подвергались справедливой критике других ученых, например, когда я вляпался в переподгонку (об этом феномене мы поговорим в четвертой главе). См. Bailey D. H., Littlefield A. K. *Does Reading Cause Later Intelligence? Accounting for Stability in Models of Change.* Child Development. 88, no. 6 (2017): 1913–21. Я даже опубликовал статью о гене-кандидате, использовав метод, по которому мы пройдемся в пятой главе. См. Ritchie S. J. et al. *Alcohol Consumption and Lifetime Change in Cognitive Ability: A Gene × Environment Interaction Study.* AGE. 36, no. 3 (2014): 9638. И почти наверняка я виновен в подверженности хайпу: в некоторых беседах с журналистами о науке я позволял себе слишком вольные формулировки или позже — все мы крепки задним умом — сожалел, что не добавил какие-то ценные замечания и не сделал важные оговорки. И я совершил ошибку, утверждая, что “опубликованы сотни рецензированных статей на эту тему”, как будто это служит индикатором правды. И что касается рецензирования: уж точно были случаи, когда я не посвящал статье, которую рецензировал, должного количества времени и мог нечаянно пропустить какие-то ошибки. Нисколько не сомневаюсь, что в будущем вскроются еще какие-нибудь мои оплошности или упущения.

Глава 1. **Как работает наука**

1 Hume D. *Of Essay-Writing // Essays: Moral, Political, and Literary*. Indianapolis: Liberty Fund, 1777.

2 Sokal A., Bricmont J. *Intellectual Impostures*. London: Profile Books, 1998, 2003.

3 Mill J. S. *On Liberty.* London: Dover Press, 1859. [Милль Дж. С. *О свободе // Антология мировой либеральной мысли (I половины XX века)*. М.: Прогресс-Традиция, 2000.]

4 Longino H. E. *Science as Social Knowledge*. Princeton: Princeton University Press, 1990. См. также Longino H. *The Social Dimensions of Scientific Knowledge.* The Stanford Encyclopedia of Philosophy. Summer 2019; Reiss J., Sprenger J. *Scientific Objectivity*. The Stanford Encyclopedia of Philosophy. Winter 2017.

5 Это утверждение я делаю под влиянием идеи специалистов по эволюционной теории Уго Мерсье и Дэна Спербера, согласно которой основная функция человеческого мышления как такового состоит в том, чтобы придумать, как лучше всего убедить других людей. См. Mercier H., Sperber D. *Why Do Humans Reason? Arguments for an Argumentative Theory*. Behavioral and Brain Sciences. 34, no. 2 (2011): 57–74.

6 McDougall-Waters J. et al. *Philosophical Transactions: 350 Years of Publishing at the Royal Society (1665–2015)*. London: Royal Society, 2015. Некоторые историки возразили бы, что первым научным журналом следует считать французское издание *Le Journal des sçavans*, вышедшее в 1665 году всего за два месяца до *Philosophical Transactions*. Однако в нем публиковались статьи по огромному числу разных ученых тем и поначалу — в основном книжные рецензии и отрывки. Тогда как *Philosophical Transactions* исходно был посвящен главным образом научным новостям и наблюдениям. Наверное, справедливее было бы признать *Le Journal des sçavans* первым академическим изданием, а *Philosophical Transactions* — первым научным. См. McCutcheon R. P. *The "Journal Des Scavans" and the "Philosophical Transactions of the Royal Society"*. Studies in Philology. 21, no. 4 (1924): 626–8; Banks D. *Thoughts on Publishing the Research Article over the Centuries*. Publications. 6, no. 1 (2018): 10.

7 David P. A. *The Historical Origins of "Open Science": An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution*. Capitalism and Society. 3, no. 2 (2008): 5.

8 Hooke R. *A Spot in One of the Belts of Jupiter.* Philosophical Transactions. 1 (1665): 3.

9 В 1900 году он был разделен на два журнала: один по математике и физическим наукам, другой по наукам биологическим. См. *https://royalsocietypublishing.org/journal/rstl.*

10 Ware M., Mabe M. *The STM Report: An Overview of Scientific and Scholarly Journal Publishing*. The Hague, Netherlands: International Association of Scientific, Technical and Medical Publishers, 2015.

11 Обратите внимание, что до середины XVIII века, когда журнал *Philosophical Transactions* стал официальным изданием Королевского общества, он издавался разными отдельными исследователями и составителями.

12 Большинство журнальных статей — они называются “экспериментальными” — сообщают о новых результатах исследований, но некоторые являются “обзорными” статьями, обобщающими все, что уже известно по какому-то конкретному научному вопросу.

13 *www.nih.gov* и *www.nsf.gov.* Похожие организации в других странах: Государственное агентство исследований и инноваций Великобритании (*UK Research and Innovation; www.ukri.org)*, Национальный фонд естественных наук Китая (*National Natural Science Foundation of China*; *www.nsfc.gov.cn/english/site_1/index.html)*, Японское общество содействия науке (*Japan Society for the Promotion of Science*; *www.jsps.go.jp/english).* См. также *https://wellcome.ac.uk/* и *www.gatesfoundation.org.*

14 Например, в некоторых научных журналах раздел “Методы” помещается в самый конец статьи, будто бы эта важнейшая информация — всего лишь второстепенное дополнение.

15 *www.sciencemag.org/site/feature/contribinfo/faq/index.xhtml#pct_faq*

16 Csiszar A. *Peer Review: Troubled from the Start*. Nature. 532, no. 7599 (2016): 306–8.

17 Цитируется по: Baldwin M. *Scientific Autonomy, Public Accountability and the Rise of “Peer Review” in the Cold War United States*. Isis. 109, no. 3 (2018): 538–58.

18 Там же.

19 *https://shitmyreviewerssay.tumblr.com*

20 Стоит отметить, что эти правила касаются научного поиска и анализа, они не имеют отношения к этическим нормам, которые все ученые также должны принимать во внимание. Эти правила, пожалуй, особенно важны для тех, кто проводит исследования на людях (или других животных), а еще для тех, кто работает с потенциально опасными технологиями или чьи эксперименты способны нанести экологический либо иной вред.

21 Merton R. K. *The Normative Structure of Science* (1942) // *The Sociology of Science: Empirical and Theoretical Investigations*. Chicago and London: University of Chicago Press, 1973.

22 Darwin Correspondence Project, Letter no. 2122, 9 July 1857. *www.darwinproject.ac.uk/letter/DCP-LETT-2122.xml*

23 Вообще Мертон назвал коллективизм “коммунизмом”, но этот термин, скажем так, имеет иные коннотации. В последующих ра-

ботах название сменилось на “коллективизм”, им я здесь и пользуюсь. См., например: Anderson M. S. et al. *Extending the Mertonian Norms: Scientists’ Subscription to Norms of Research*. Journal of Higher Education. 81, no. 3 (2010): 366–93.

24 Мертон упоминает нелюдимого Генри Кавендиша, физика и химика, жившего в XVIII веке, как исторического нарушителя этой нормы: тот скрыл от мира множество своих важных экспериментов и теорий из чистой застенчивости, и лишь много позже его смерти эти результаты были получены заново.

25 Best N. W. *Lavoisier’s “Reflections on Phlogiston” I: Against Phlogiston Theory.* Foundations of Chemistry. 17, no. 2 (2015): 137–51.

26 Dawkins R. *The God Delusion.* London: Bantam Books, 2006. [Докинз Р. *Бог как иллюзия*. М.: КоЛибри, 2018.]

27 Planck M. *Scientific Autobiography and Other Papers.* London: Williams & Norgate, Ltd., 1949. [См., например: Планк М. *Научная автобиография.* УФН. 64, № 4 (1958): 625–37.]

28 Popper K. *The Logic of Scientific Discovery*. London & New York: Routledge Classics, 1959/2002. [Поппер К. *Логика научного исследования* // Поппер К. *Логика и рост научного знания*. М.: Прогресс, 1983.]

29 Подобное повторение опыта, хоть и проведенное лет на сто позже, изображено — в драматическом свете — на картине Джозефа Райта (Райта из Дерби) “Эксперимент с птицей в воздушном насосе”, сейчас находящейся в Лондонской национальной галерее.

30 Boyle R. *The New Experiments Physico-Mechanical, Touching the Spring of the Air and Its Effects*. London: Miles Flesher, 1682. Цит. по: Shapin S., Schaffer S. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life.* Princeton: Princeton University Press, 1985.

31 Shapin S., Schaffer S. *Leviathan*.

Глава 2. **Кризис воспроизводимости**

1 Nosek B. A. et al. *Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability*. Perspectives on Psychological Science. 7, no. 6 (2012): 615–31.

2 Kahneman D. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011. [Канеман Д. *Думай медленно… решай быстро*. М.: АСТ, 2013.]

3 Neely J. *Semantic Priming Effects in Visual Word Recognition: A Selective Review of Current Findings and Theories // Basic Processes in Reading: Visual Word Recognition*. Abingdon: Routledge, 2012.

4 Zhong C. B., Liljenquist K. *Washing Away Your Sins: Threatened Morality and Physical Cleansing*. Science. 313, no. 5792 (2006): 1451–2.

5 Vohs K. D. et al. *The Psychological Consequences of Money.* Science. 314, no. 5802 (2006): 1154–6.
6 Там же.
7 Kahneman D. *Thinking, Fast and Slow*.
8 Насколько мне известно, термин происходит из статьи Пашлера и Вагенмэйкерса, которые не использовали словосочетание “кризис воспроизводимости” впрямую, но говорили о “кризисе доверия” в психологических исследованиях после серии неудавшихся повторений. Нельсон, Симмонс и Саймонсон обсуждали причины возникновения кризиса. См. Pashler H., Wagenmakers E.-J. *Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?* Perspectives on Psychological Science. 7, no. 6 (2012): 528–30; Nelson L. D. et al. *Psychology's Renaissance*. Annual Review of Psychology. 69, no. 1 (2018): 511–34.
9 Bargh J. A. et al. *Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action*. Journal of Personality and Social Psychology. 71, no. 2 (1996): 230–44.
10 Количество цитирований (точное число — 5208) дано по состоянию на январь 2020 года в соответствии с *Google Scholar*.
11 Doyen S. et al. *Behavioral Priming: It's All in the Mind, but Whose Mind?* PLOS ONE. 7, no. 1 (2012): e29081.
12 Earp B. D. et al. *Out, Damned Spot: Can the “Macbeth Effect” Be Replicated?* Basic and Applied Social Psychology. 36, no. 1 (2014): 91–8. Эффект денежного прайминга: Klein R. A. et al. *Investigating Variation in Replicability: A “Many Labs” Replication Project*. Social Psychology. 45, no. 3 (2014): 142–52.
13 Исходное исследование: Williams L. E., Bargh J. A. *Keeping One's Distance: The Influence of Spatial Distance Cues on Affect and Evaluation*. Psychological Science. 19, no. 3 (2008): 302–8. Повторение: Pashler H. et al. *Priming of Social Distance? Failure to Replicate Effects on Social and Food Judgments*. PLOS ONE. 7, no. 8 (2012): e42510.
14 Исходное исследование: Zarkadi T., Schnall S. *“Black and White” Thinking: Visual Contrast Polarizes Moral Judgment*. Journal of Experimental Social Psychology. 49, no. 3 (2013): 355–9. Повторение: IJzerman H., Laine P.-J. *Does Background Color Affect Moral Judgment? Three Pre-Registered Replications of Zarkadi and Schnall's (2012) Study 1.* Preprint, PsyArXiv (30 July 2018).
15 Для прайминга “отвращением” экспериментаторы часто делали так, чтобы в помещении дурно пахло. Поэтому исследования на эту тему особенно примечательны: во многих статьях психологам приходилось с непроницаемым видом разглагольствовать об эффектах “спрея с запахом кишечных газов”, а в одной работе авторы невозмутимо обсуждали “патентованный одорант под названием «Жид-

кая задница»". О средстве "Жидкая задница" см. Adams T. G. et al. *The Effects of Cognitive and Affective Priming on Law of Contagion Appraisals.* Journal of Experimental Psychopathology. 3, no. 3 (2012): 473. Обзор этого направления исследований: Landy J. F., Goodwin G. P. *Does Incidental Disgust Amplify Moral Judgment? A Meta-Analytic Review of Experimental Evidence*. Perspectives on Psychological Science. 10, no. 4 (2015): 518–36.

16 McCook A. *"I Placed Too Much Faith in Underpowered Studies": Nobel Prize Winner Admits Mistakes*. Retraction Watch. 20 Feb. 2017. Еще Канеман написал открытое письмо, адресованное социальным психологам, в котором сообщил, что узрел "угрозу катастрофы", и призвал их изменить свой подход к исследованиям; см. тут: *https://go.nature.com/2T7A2NV.*

17 Carney D. R. et al. *Power Posing: Brief Nonverbal Displays Affect Neuroendocrine Levels and Risk Tolerance.* Psychological Science. 21, no. 10 (2010): 1363–8.

18 На момент написания этой книги в феврале 2020 года общее количество просмотров на сайте TED составляло 56 миллионов, а на *YouTube* — еще 17,6 миллиона. Выступление исходно называлось "Язык тела формирует вашу личность", но затем, когда грянул кризис воспроизводимости, оно было переименовано и стало называться "Язык тела может формировать вашу личность". Cuddy A. *Your Body Language May Shape Who You Are.* TEDGlobal 2012. June 2012.

19 Cuddy A. J. C. *Presence: Bringing Your Boldest Self to Your Biggest Challenges.* New York: Little, Brown and Company, 2015. Цитаты взяты с издательского сайта: *www.littlebrown.com/titles/amy-cuddy/presence/9780316256575.*

20 Khaleeli H. *A Body Language Lesson Gone Wrong: Why is George Osborne Standing like Beyoncé?* The Guardian. 7 Oct. 2015.

21 Ranehill E. et al. *Assessing the Robustness of Power Posing: No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women*. Psychological Science. 26, no. 5 (2015): 653–6. С тех пор дебаты по поводу поз силы только набирали обороты. Авторы обзора 2017 года заключили, что эффекты поз силы — это "предположения, на данный момент не имеющие эмпирических обоснований". См. Simmons J. P., Simonsohn U. *Power Posing: P-Curving the Evidence*. Psychological Science. 28, no. 5 (2017): 687–93. Кадди парировала своим собственным обзором, где отмечался-таки общий эффект, хотя позже и было показано, что — наряду с другими проблемами подобных исследований — большинство результатов в статьях, на которые она ссылалась, объяснялись, вероятно, отрицательным эффектом ссутуливания, нежели чем положительным воздействием поз силы. См. Cuddy A. J. C. et al. *P-Curving a More Comprehen-*

sive Body of Research on Postural Feedback Reveals Clear Evidential Value for Power-Posing Effects: Reply to Simmons and Simonsohn (2017). Psychological Science. 29, no. 4 (2018): 656–66. Про ссутуливание см. Credé M. *A Negative Effect of a Contractive Pose is not Evidence for the Positive Effect of an Expansive Pose: Commentary on Cuddy, Schultz, and Fosse (2018).* SSRN. 2018.

22 Zimbardo P. *The Lucifer Effect: How Good People Turn Evil*. London: Rider, 2007. [Зимбардо Ф. *Эффект Люцифера. Почему хорошие люди превращаются в злодеев*. М.: АНФ, 2013.]

23 Milgram S. *Behavioral Study of Obedience.* Journal of Abnormal and Social Psychology. 67, no. 4 (1963): 371–8. Эксперименты Милгрэма тоже подвергались справедливой критике. Чем сильнее участники верили, что действительно бьют “учеников” током, тем с меньшей вероятностью повышали разряд. Об этом см., например: Perry G. et al. *Credibility and Incredulity in Milgram’s Obedience Experiments: A Reanalysis of an Unpublished Test*. Social Psychology Quarterly. 83, no. 1 (2020): 88–106.

24 Zimbardo P. *Our inner heroes could stop another Abu Ghraib*. The Guardian. 29 Feb. 2008.

25 Fromm E. *The Anatomy of Human Destructiveness*. New York: Holt, Rinehart and Winston, 1975. [Фромм Э. *Анатомия человеческой деструктивности.* М.: АСТ, 2004.]

26 Le Texier T. *Debunking the Stanford Prison Experiment.* American Psychologist. 74, no. 7 (2019): 823–39.

27 Дебаты продолжаются, и Зимбардо ответил на критику. См., например: Zimbardo P. *Philip Zimbardo’s Response to Recent Criticisms of the Stanford Prison Experiment*. 23 June 2018. См. также ответ Ле Тексье на более позднюю версию заявления Зимбардо (на момент написания этой книги — еще не опубликованную): Le Texier T. *The SPE Remains Debunked: A Reply to Zimbardo and Haney (2020)*. Preprint, PsyArXiv (24 Jan. 2020).

28 Open Science Collaboration. *Estimating the Reproducibility of Psychological Science*. Science. 349, no. 6251 (2015): aac4716.

29 Camerer C. F. et al. *Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015*. Nature Human Behaviour. 2, no. 9 (2018): 637–44.

30 Последнее число соответствует шести удачным попыткам воспроизвести шестнадцать исследований. Ebersole C. R. et al. *Many Labs 3: Evaluating Participant Pool Quality across the Academic Semester via Replication*. Journal of Experimental Social Psychology. 67 (2016): 68–82.

31 Тут некоторые критики могут возразить, что я сам попал в вырытую другим яму. Я подчеркивал важность надежных результатов, однако, заявляя, что разразился кризис воспроизводимости, полагаюсь

на попытки воспроизвести разные исследования, которые не являются репрезентативной выборкой из всей научной литературы. Вывод о том, что лишь "около половины" опубликованных результатов воспроизводятся, возможно, нельзя обобщать на всю науку. Такой аргумент был приведен в критической заметке к одному из исследований, посвященных повторению других работ: Gilbert D. T. et al. *Comment on "Estimating the Reproducibility of Psychological Science"*. Science. 351, no. 6277 (2016): 1037. Хотя я не согласен со многими приведенными там аргументами (некоторые причины скепсиса описаны здесь: Lakens D. *The Statistical Conclusions in Gilbert et al (2016) Are Completely Invalid.* The 20 % Statistician. 6 March 2016), замечание насчет репрезентативности справедливо. Мы все еще плохо понимаем, какая именно часть результатов по всем научным направлениям воспроизводима, даже в областях вроде психологии, где были предприняты эти масштабные попытки повторить прежние результаты, — правда может оказаться пригляднее, чем показывают те исследования, или же наоборот. Но сам факт, что мы этого не знаем, — а также то, что столь много громких, нашумевших открытий рассыпалось в прах при более внимательном рассмотрении, — является, я убежден, поводом для немалого беспокойства. Ответы на другие критические замечания, оспаривающие идею, что разразился кризис, см. тут: Pashler H., Harris C. R. *Is the Replicability Crisis Overblown? Three Arguments Examined*. Perspectives on Psychological Science. 7, no. 6 (2012): 531–6.

32 Bird A. *Understanding the Replication Crisis as a Base Rate Fallacy*. British Journal for the Philosophy of Science. 13 Aug. 2018.

33 Разумеется, авторы исходной работы (те, чьи результаты не воспроизвелись) зачастую заявляли, что изменения были на самом деле значительными и серьезно испортили эксперимент. Каждый случай следует разбирать отдельно, однако подобный аргумент наводит на мысль о некой предвзятости.

34 Другая область, где все неплохо, — психология личности. Психолог Кристофер Сото провел масштабную работу по воспроизведению результатов исследований личности — корреляций личностных качеств, оцениваемых по опросникам, с такими показателями, как удовлетворенность жизнью и романтическими отношениями, религиозные и политические взгляды и карьерный успех. Доля успешно воспроизведенных результатов получилась 87 %, что весьма достойно по сравнению с другими областями, которые мы обсуждали. Soto C. J. *How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project*. Psychological Science. 30, no. 5 (2019): 711–27.

35 Camerer C. F. et al. *Evaluating Replicability of Laboratory Experiments in Economics.* Science. 351, no. 6280 (2016): 1433–6.

36 Turner B. O. et al. *Small Sample Sizes Reduce the Replicability of Task-Based fMRI Studies.* Communications Biology. 1, no. 1 (2018): 62.

37 Eklund A. et al. *Cluster Failure: Why fMRI Inferences for Spatial Extent Have Inflated False-Positive Rates*. Proceedings of the National Academy of Sciences. 113, no. 28 (2016): 7900–5; Eklund A. et al. *Cluster Failure Revisited: Impact of First Level Design and Physiological Noise on Cluster False Positive Rates*. Human Brain Mapping. 40, no. 7 (2019): 2017–32.

38 Lord K. A. et al. *The History of Farm Foxes Undermines the Animal Domestication Syndrome*. Trends in Ecology & Evolution. 35, no. 2 (2020): 125–36.

39 Зебровые амадины: Wang D. et al. *Irreproducible Text-Book "Knowledge": The Effects of Color Bands on Zebra Finch Fitness.* Evolution. 72, no. 4 (2018): 961–76. См. также Law Y.-H. *Replication Failures Highlight Biases in Ecology and Evolution Science.* The Scientist. 31 July 2018. Воробьи: Sánchez-Tójar A. et al. *Meta-analysis challenges a textbook example of status signalling and demonstrates publication bias*. eLife. 7 (2008): e37385. Обыкновенные лазоревки: Parker T. H. *What Do We Really Know about the Signalling Role of Plumage Colour in Blue Tits? A Case Study of Impediments to Progress in Evolutionary Biology*. Biological Reviews. 88, no. 3 (2013): 511–36.

40 Clark T. D. et al. *Ocean Acidification Does Not Impair the Behaviour of Coral Reef Fishes*. Nature. 577, no. 7790 (2020): 370–5. См. также Enserink M. *Analysis Challenges Slew of Studies Claiming Ocean Acidification Alters Fish Behavior*. Science. 8 Jan. 2020. Как отмечается в этой второй статье, из того, что поведение рыб, похоже, не меняется, не следует, что нам нужно перестать беспокоиться о закислении океана, которое вызывает много других пагубных явлений. Позднее авторов исходных статей (двадцати двух штук) о закислении океана и поведении рыб обвинили в научном мошенничестве. См. Enserink M. *Does Ocean Acidification Alter Fish Behavior? Fraud Allegations Create a Sea of Doubt.* Science. 6 May 2021.

41 *www.orgsyn.org/instructions.aspx.* См. также Chawla D. S. *Taking on Chemistry's Reproducibility Problem*. Chemistry World. 20 March 2017.

42 Поиск нужных статей проводился таким образом, что исследования, которые открыто не объявляли себя попытками воспроизвести предыдущие работы, могли оказаться неучтенными, поэтому итоговый процент, вероятно, на самом деле чуточку больше. Экономика: Mueller-Langer F. et al. *Replication Studies in Economics — How Many and Which Papers Are Chosen for Replication and Why?* Research Policy. 48, no. 1 (2019): 62–83. Психология: Makel M. C. et al. *Replications in Psychology Research: How Often Do They Really Occur?* Perspectives on Psychological Science. 7, no. 6 (2012): 537–42. Также обращаю ваше внимание: по поводу того, что

считать попыткой воспроизвести исследование, ведутся споры. Некоторые ученые провели множество "содержательных" повторений, в целом похожих на исходное исследование, но в деталях иногда от него отличающихся. Это по-своему интересно, но это не "прямое" повторение, когда именно та же, насколько только возможно, работа проводится сызнова. Вот такого рода исследований-повторений и не хватает. См. Schmidt S. *Shall We Really Do It Again? The Powerful Concept of Replication is Neglected in the Social Sciences*. Review of General Psychology. 13, no. 2 (2009): 90–100.

43 Chang A. C., Li P. *Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals say "Usually Not"*. Finance and Economics Discussion Series. 2015, no. 83 (2015): 1–26. Washington: Board of Governors of the Federal Reserve System. Подробный обзор проблемы воспроизводимости в экономике: Christensen G., Miguel E. *Transparency, Reproducibility, and the Credibility of Economics Research*. Working Paper no. 22989. National Bureau of Economic Research. 2016.

44 Konkol M. et al. *Computational Reproducibility in Geoscientific Papers: Insights from a Series of Studies with Geoscientists and a Reproduction Study*. International Journal of Geographical Information Science. 33, no. 2 (2019): 408–29.

45 И даже хуже: из этих семи статей в целых шести методы избыточны по сравнению с гораздо более простыми методами, которые были известны за много лет до того, как создавались эти новые алгоритмы. Dacrema M. F. et al. *Are We Really Making Much Progress? A Worrying Analysis of Recent Neural Recommendation Approaches*. RecSys 2019. Copenhagen, Denmark. См. также отчет по компьютерным наукам, который дает понять, что у новых исследователей не получается воспроизвести результаты применения нескольких классических алгоритмов — а это своего рода бомба замедленного действия, поскольку "молодые ученые не хотят выглядеть хулителями старших коллег", публикуя статьи о провалившихся попытках воспроизвести действие алгоритмов, которые были разработаны старшими коллегами и на которых держится их репутация. Hutson M. *Artificial Intelligence Faces Reproducibility Crisis*. Science. 359, no. 6377 (2018): 725–6.

46 Begley C. G., Ellis L. M. *Raise Standards for Preclinical Cancer Research*. Nature. 483, no. 7391 (2012): 531–3.

47 Prinz F. et al. *Believe It or Not: How Much Can We Rely on Published Data on Potential Drug Targets?* Nature Reviews Drug Discovery. 10 (2011): 712. Обратите внимание, что приведенная в статье сотрудниками *Bayer* диаграмма включает только 70 % исследований рака — остальные 30 % связаны с женским здоровьем или сердечно-сосудистой системой.

48 Wong C. H. et al. *Estimation of Clinical Trial Success Rates and Related Parameters*. Biostatistics. 20, no. 2 (2019): 273–86. Из всех раз-

нообразных лекарств доля тех, что добираются от доклинических испытаний до применения на людях, оценивается в данной работе в 13,8 %, так что с исследованиями рака все особенно плохо.

49 Nosek B. A., Errington T. M. *Reproducibility in Cancer Biology: Making Sense of Replications*. eLife. 6 (2017): e23383. В названии "Проект по проверке воспроизводимости: биология рака" термин "воспроизводимость" используется в том же смысле, в каком я использую слово "сходимость" (то есть попытка получить те же результаты на другой выборке). Я выбрал определения для этой книги таким образом, чтобы отразить сложившийся консенсус, однако стоит понимать, что не все придерживаются той же терминологии.

50 Repass J. et al. *Replication Study: Fusobacterium Nucleatum Infection is Prevalent in Human Colorectal Carcinoma*. eLife. 7 (2018): e25801.

51 Errington T. *Reproducibility Project: Cancer Biology — Barriers to Replicability in the Process of Research*. 2019. *https://osf.io/x9p5s/*

52 Baker M., Dolgin E. *Cancer Reproducibility Project Releases First Results*. Nature. 541, no. 7637 (2017): 269–70; Engber D. *Cancer Research Is Broken*. Slate. 19 April 2016.

53 Errington T. *Reproducibility Project* (см. слайд 11).

54 Kaiser J. *The Cancer Test*. Science. 348, no. 6242 (2015): 1411–3.

55 Iqbal S. A. et al. *Reproducible Research Practices and Transparency across the Biomedical Literature*. PLOS Biology. 14, no. 1 (2016): e1002333. Обратите внимание, что в полную выборку вошло 441 исследование, но только 268 из них содержали эмпирические данные.

56 Vasilevsky N. A. et al. *On the Reproducibility of Science: Unique Identification of Research Resources in the Biomedical Literature.* PeerJ. 1 (2013): e148. Проблемы, касающиеся недостаточно подробного написания статей, выходят за пределы биомедицины. О политологии, например, см. Wuttke A. *Why Too Many Political Science Findings Cannot Be Trusted and What We Can Do About It: A Review of Meta-Scientific Research and a Call for Academic Reform.* Politische Vierteljahresschrift. 60, no. 1 (2019): 1–19. Об экологии см. Parker T. H. et al. *Transparency in Ecology and Evolution: Real Problems, Real Solutions*. Trends in Ecology & Evolution. 31, no. 9 (2016): 711–9.

57 Kaiser J. *Plan to Replicate 50 High-Impact Cancer Papers Shrinks to Just 18.* Science. 31 July 2018. Обратите внимание, что в примечании 51 к этой главе (Errington T. *Reproducibility Project*) обсуждается пятьдесят одно исследование, а не пятьдесят.

58 Все о "Проекте по проверке воспроизводимости: биология рака" см. здесь: *https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology.*

59 Prasad V. K., Cifu A. S. *Ending Medical Reversal: Improving Outcomes, Saving Lives*. Baltimore: Johns Hopkins University Press, 2015.

60 Lang J. *Awakening*. The Atlantic. Feb. 2013.

61 Avidan M. S. et al. *Anesthesia Awareness and the Bispectral Index*. New England Journal of Medicine. 358, no. 11 (2008): 1097.
62 Herrera-Perez D. et al. *A Comprehensive Review of Randomized Clinical Trials in Three Medical Journals Reveals 396 Medical Reversals*. eLife. 8 (2019): e45183. Это было продолжением похожего исследования тех же авторов, когда обнаружилось 146 "медицинских реверсий": Prasad V. et al. A *Decade of Reversal: An Analysis of 146 Contradicted Medical Practices.* Mayo Clinic Proceedings. 88, no. 8 (2013): 790–8.
63 Barrett J. F. R. et al. *A Randomized Trial of Planned Cesarean or Vaginal Delivery for Twin Pregnancy*. New England Journal of Medicine. 369, no. 14 (2013): 1295–1305.
64 Du Toit G. et al. *Randomized Trial of Peanut Consumption in Infants at Risk for Peanut Allergy*. New England Journal of Medicine. 372, no. 9 (2015): 803–13.
65 Kim F. et al. *Effect of Prehospital Induction of Mild Hypothermia on Survival and Neurological Status Among Adults with Cardiac Arrest: A Randomized Clinical Trial*. JAMA. 311, no. 1 (2014): 45–52.
66 AVERT Collaboration. *Efficacy and Safety of Very Early Mobilisation within 24 h of Stroke Onset: A Randomised Controlled Trial*. Lancet. 386, no. 9988 (2015): 46–55.
67 Baharoglu M. I. et al. *Platelet Transfusion versus Standard Care after Acute Stroke Due to Spontaneous Cerebral Haemorrhage Associated with Antiplatelet Therapy (PATCH): A Randomised, Open-Label, Phase 3 Trial*. Lancet. 387, no. 10038 (2016): 2605–13.
68 Villas Boas P. J. F. et al. *Systematic Reviews Showed Insufficient Evidence for Clinical Practice in 2004: What about in 2011? The Next Appeal for the Evidence-Based Medicine Age*. Journal of Evaluation in Clinical Practice. 19, no. 4 (2013): 633–7.
69 Freedman L. P. et al. *The Economics of Reproducibility in Preclinical Research*. PLOS Biology. 13, no. 6 (2015): e1002165.
70 Chalmers I., Glasziou P. *Avoidable Waste in the Production and Reporting of Research Evidence*. Lancet. 374, no. 9683 (2009): 86–9. См. также Macleod M. R. et al. *Biomedical Research: Increasing Value, Reducing Waste*. Lancet. 383, no. 9912 (2014): 101–4.
71 Baker M. *1,500 Scientists Lift the Lid on Reproducibility*. Nature. 533, no. 7604 (2016): 452–4.
72 Ioannidis J. P. A. *Why Most Published Research Findings Are False*. PLOS Medicine. 2, no. 8 (2005): e124.
73 Количество цитирований — из *Google Scholar*.
74 Критический разбор статьи Иоаннидиса см., например, тут: Leek J. T., Jager L. R. *Is Most Published Research Really False?* Annual Review of Statistics and Its Application. 4, no. 1 (2017): 109–22.

Глава 3. **Мошенничество**

1 MacDonald N. *Maxims and Moral Reflections*. New York: Collins & Hannay, 1827.

2 Кохлеарные импланты: Williams V. *Baby Hears for First Time with Cochlear Implants*. Mayo Clinic News Network. 13 Nov. 2018; *www.youtube.com/watch?v=lfbQbQ-NWB0.* Катаракты: *Two Blind Sisters See for the First Time*. National Geographic. 26 Sept. 2014; *www.youtube.com/watch?v=EltIpB4EtYU.* Протезы: Smith V. *Video Of Rick Clement Walking On New Legs Goes Viral*. Forces Network. 23 July 2015; *www.forces.net/services/tri-service/video-rick-clement-walking-new-legs-goes-viral;* а также: *Boy, 5, given Prosthetic Arm That Lets Him Hug Brother*. BBC News. 14 Dec. 2019; *www.bbc.co.uk/news/uk-wales-50762563.*

3 Соединение — это процедура под названием “анастомоз”, которая часто дает хорошие результаты. Развитие хирургических методик в начале и середине XX века позволило сделать так, чтобы анастомоз оставался возможен после удаления все более длинных частей трахеи. Эти усилия в конце концов достигли своего предела: если приходилось удалять больше половины длины трахеи, что происходит в тех случаях, когда опухоли особенно сильно разрослись, накладывать анастомоз было уже нельзя.

4 Grillo H. C. *Tracheal Replacement: A Critical Review.* The Annals of Thoracic Surgery. 73, no. 6 (2002): 1995–2004.

5 Macchiarini P. et al. *Clinical Transplantation of a Tissue-Engineered Airway*. Lancet. 372, no. 9655 (2008): 2023–30.

6 *First Successful Transplantation of a Synthetic Tissue Engineered Windpipe*. Karolinska Institute. News release. 29 July 2011.

7 Jungebluth P. et al. *Tracheobronchial Transplantation with a Stem-Cell-Seeded Bioartificial Nanocomposite: A Proof-of-Concept Study*. Lancet. 378, no. 9808 (2011): 1997–2004.

8 Berggren C., Karabag S. F. *Scientific Misconduct at an Elite Medical Institute: The Role of Competing Institutional Logics and Fragmented Control*. Research Policy. 48, no. 2 (2019): 428–43.

9 Там же.

10 Huss M. S. *The Macchiarini Case: Timeline*. Karolinska Institute. 26 June 2018.

11 *Macchiarini’s Seventh Transplant Patient Dies*. The Local. 20 March 2017.

12 Цитируется по: Berggren C., Karabag S. F. *Scientific Misconduct.* Оригинальный источник: Wahlström J. *Den Bortglömda Patienten.* Filter. 18 May 2016.

13 Kremer W. *Paolo Macchiarini: A Surgeon’s Downfall*. BBC News Magazine. 10 Sept. 2016.

14 Хоть она и была здорова, после автомобильной аварии ей сделали трахеостомию: чтобы говорить, ей приходилось закрывать отверстие рукой. Она надеялась, что после операции сможет петь своему сыну. Elliott C. *Knifed with a Smile*. New York Review of Books. 5 April 2018.

15 Британский пациент, которого до того оперировал Маккиарини, получил синтетическую трахею в стиле итальянского хирурга от его коллег в Лондоне в 2011-м, но в следующем году умер. Kremer W. *Paolo Macchiarini*.

16 Сердце разрывается, когда сравниваешь статьи в одном и том же канадском издании, вышедшие до и после смерти девочки: *"We Feel like She's Reborn": Toddler Born without Windpipe Gets New One Grown from Her Own Stem Cells*. National Post. 30 April 2013; Brean J. *Swashbuckling Surgeon's Collapsing Reputation Threatens Canadian Girl's Legacy as "Pioneer" Patient*. National Post. 18 Feb. 2016.

17 Herold E. *A Star Surgeon Left a Trail of Dead Patients — and His Whistleblowers were Punished*. Leapsmag. 8 Oct. 2018.

18 Отчет размещен на сайте *Retraction Watch*: *http://retractionwatch.com/wp-content/uploads/2015/05/Translation-investigation.doc* (см. с. 36).

19 Cyranoski D. *Artificial-Windpipe Surgeon Committed Misconduct*. Nature. 521, no. 7553 (2015): 406–7. См. также McCook A. *Misconduct Found in 7 Papers by Macchiarini, Says English Write-up of Investigation*. Retraction Watch. 28 May 2015.

20 Kremer W. *Paolo Macchiarini*.

21 *Paolo Macchiarini Is Not Guilty of Scientific Misconduct.* Lancet. 386, no. 9997 (2015): 932.

22 Ciralsky A. *The Celebrity Surgeon Who Used Love, Money, and the Pope to Scam an NBC News Producer*. Vanity Fair. 5 Jan. 2016.

23 Там же.

24 В ходе расследования сотрудники *Vanity Fair* также заподозрили, что Маккиарини выдумал многие из своих профессиональных умений и мест работы, которые указывал в резюме, и что факты, всплывшие в ходе итальянского расследования, так и не были обнародованы.

25 Kremer W. *Paolo Macchiarini*.

26 Huss M. S. *The Macchiarini Case*. См. также Cyranoski D. *Nobel Official Resigns over Karolinska Surgeon Controversy*. Nature. 8 Feb. 2016.

27 Маккиарини действовал не один: в отчете Каролинского института в недобросовестной научной практике также обвинялись несколько его соавторов.

28 *The Final Verdict on Paolo Macchiarini: Guilty of Misconduct*. Lancet. 392, no. 10141 (2018): 2.

29 *Seven Researchers Responsible for Scientific Misconduct in Macchiarini Case*. Karolinska Institute. 28 June 2015.

30 Warren M. *Disgraced Surgeon is Still Publishing on Stem Cell Therapies*. Science. 27 April 2018.

31 Zhuravleva M. et al. *In Vitro Assessment of Electrospun Polyamide-6 Scaffolds for Esophageal Tissue Engineering.* Journal of Biomedical Materials Research Part B: Applied Biomaterials. 107, no. 2 (2019): 253–68.

32 Astakhova A. *Superstar Surgeon Fired, Again, This Time in Russia*. Science. 16 May 2017.

33 *Investigation Concerning Surgeries Resumed after Review*. Swedish Prosecution Authority. 11 Dec. 2018.

34 Herold E. *A Star Surgeon Left a Trail of Dead Patients.*

35 Berggren C., Karabag S. F. *Scientific Misconduct.*

36 Еще Саммерлин фальсифицировал результаты якобы совершенной им трансплантации роговицы у кроликов. Brody J. E. *Inquiry at Cancer Center Finds Fraud in Research*. New York Times. 25 May 1974. См. также Medawar P. *The Strange Case of the Spotted Mice // The Strange Case of the Spotted Mice: And Other Classic Essays on Science*. Oxford: Oxford University Press, 1996.

37 Pande P. G. et al. *Toxoplasma from the Eggs of the Domestic Fowl (Gallus gallus)*. Science. 133, no. 3453 (1961): 648.

38 DuShane G. et al. *An Unfortunate Event*. Science. 134, no. 3483 (1961): 945–6. См. также Kavanau J. L., Norris K. S. *Letter to the Editor*. Science. 136, no. 3511 (1962): 199; Wade N. B., Broad W. *Betrayers of the Truth: Fraud and Deceit in the Halls of Science.* New York: Simon & Schuster, 1982.

39 Кличка *Snuppy* складывается из аббревиатуры SNU (*Seoul National University*, Сеульский национальный университет) и слова *puppy* ("щенок"). Достижение было особенно впечатляющим, поскольку у собак по сравнению с другими млекопитающими яйцеклетки довольно хрупкие и нежные. Хотя к 2005 году уже клонировали таких животных, как овца, кошка, свинья и лошадь, никому до Хвана не удавалось клонировать собаку. Lee B. C. et al. *Dogs Cloned from Adult Somatic Cells*. Nature. 436, no. 7051 (2005): 641.

40 Park J. et al. *The Korean Press and Hwang's Fraud.* Public Understanding of Science. 18, no. 6 (2009): 653–69.

41 Saunders R., Savulescu J. *Research Ethics and Lessons from Hwanggate: What Can We Learn from the Korean Cloning Fraud?* Journal of Medical Ethics. 34, no. 3 (2008): 214–21.

42 Holden C. *Bank on These Stamps*. Science. 308, no. 5729 (2005): 1738a.

43 "Величайший ученый": Kim J., Park K. *Ethical Modernization: Research Misconduct and Research Ethics Reforms in Korea Following the Hwang Affair*. Science and Engineering Ethics. 19, no. 2 (2013): 358. Пожертвование яйцеклеток: Saunders R., Savulescu J. *Research Ethics.*

44 Couzin J. *STEM CELLS: …And How the Problems Eluded Peer Reviewers and Editors*. Science. 311, no. 5757 (2006): 23–4; Rossner M. *Hwang Case Review Committee Misses the Mark*. Journal of Cell Biology. 176, no. 2 (2007): 131–2.

45 Saunders R., Savulescu J. *Research Ethics.*

46 Kim J., Park K. *Ethical Modernization.*

47 Там же.

48 Chong S., Normile D. *STEM CELLS: How Young Korean Researchers Helped Unearth a Scandal*. Science. 311, no. 5757 (2006): 22–5.

49 Ahn M.-Y., Normile D. *Korean Supreme Court Upholds Disgraced Cloner's Criminal Sentence*. Science. 27 Feb. 2014.

50 Я уверен, что Снаппи, в отличие от мужчины, клонировавшего его, был чудесным. Kim M. J. et al. *Birth of Clones of the World's First Cloned Dog*. Scientific Reports. 7, no. 1 (2017): 15235.

51 Я испытываю совершенно иррациональное чувство гордости из-за того, что эту лабораторную методику, используемую теперь повсеместно, Саузерн изобрел в моей альма-матер — Эдинбургском университете. Southern E. M. *Detection of Specific Sequences among DNA Fragments Separated by Gel Electrophoresis*. Journal of Molecular Biology. 98, no. 3 (1975): 503–17.

52 Полностью саузерн-блоттинг выглядит примерно следующим образом. Берется молекула ДНК, разрезается ферментами на более мелкие фрагменты, которые затем прогоняются через специальный гель под действием электрического тока (этот процесс называется электрофорезом). Фрагменты разного размера движутся через гель с разной скоростью, поэтому после подачи тока в течение некоторого времени о размере фрагмента можно судить по его положению в геле. После электрофореза с помощью щелочного раствора двойная спираль ДНК расщепляется на отдельные нити (это освобождает место для зонда, как мы сейчас увидим). Затем содержимое геля переносится на специальную мембрану (обычно нитроцеллюлозную, но используется также и нейлоновая), на которую добавляется другая, изначально имевшаяся нить ДНК (тот самый "зонд"), предварительно радиоактивно помеченная. Этот зонд связывается с соответствующими участками ДНК, перенесенной на мембрану из геля, и тогда с помощью рентгеновской пленки можно увидеть, где связывание произошло (а значит, например, какие гены присутствуют в вашем образце ДНК). В итоге вы видите различные участки ДНК, выстроенные размытыми полосками разного размера и интенсивности. Вместо радиоактивных меток при блоттинге также можно использовать цветные красители.

53 С помощью нозерн-блоттинга детектируют РНК, а с помощью вестерн-блоттинга — белки. Еще есть истерн-блоттинг (для обнаружения модификаций белков) и даже фар-истерн-блоттинг

(far-eastern). Последний метод назван так потому, что имеет отношение к истерн-блоттингу, но разработан в Японии.

54 Obokata H. et al. *Stimulus-Triggered Fate Conversion of Somatic Cells into Pluripotency*. Nature. 505, no. 7485 (2014): 641–7; Obokata H. et al. *Bidirectional Developmental Potential in Reprogrammed Cells with Acquired Pluripotency*. Nature. 505, no. 7485 (2014): 676–80.

55 Это некоторое упрощение: индуцированные стволовые клетки имеют несколько иные свойства по сравнению с эмбриональными, что может в итоге оказаться весьма важно в медицинском контексте — это покажут лишь дальнейшие исследования. К слову, процесс получения индуцированных плюрипотентных стволовых клеток еще не был открыт в 2004–2005 годах, вот почему Хван сосредоточился на работе со стволовыми клетками эмбрионов.

56 Nobel Media. *The Nobel Prize in Physiology or Medicine 2012*. Oct. 2012. *www.nobelprize.org/prizes/medicine/2012/summary/*

57 Как писала газета *The Japan Times*, продажи каппоги резко выросли, когда Обоката прославилась. Hooper R. *Stem-Cell Leap Defied Japanese Norms*. Japan Times. 14 Feb. 2014.

58 Ishii S. et al. *Report on* STAP *Cell Research Paper Investigation*. 31 March 2014.

59 “Красный шрифт — для провальных результатов”: *https://ipscell.com/stap-new-data.* “Все они оказывались красными”: Meskus M. et al. *Research Misconduct in the Age of Open Science: The Case of* STAP *Stem Cells*. Science as Culture. 27, no. 1 (2018): 1–23. Еще эта статья дает интересную тему для обсуждения: как для развенчания методики STAP использовался интернет — анонимные комментарии к поддельным изображениям и блог, каталогизирующий попытки воспроизвести результаты.

60 Gallagher J. *Stem Cell Scandal Scientist Haruko Obokata Resigns*. BBC News. 19 Dec. 2014.

61 Katsura I. et al. *Report on* STAP *Cell Research Paper Investigation*. 25 Dec. 2014; Kameda M. “STAP *Cells” Claimed by Obokata Were Likely Embryonic Stem Cells*. Japan Times. 26 Dec. 2014.

62 Cyranoski D. *Collateral Damage: How One Misconduct Case Brought a Biology Institute to Its Knees*. Nature. 520, no. 7549 (2015): 600–3.

63 Cyranoski D. *Stem-Cell Pioneer Blamed Media “Bashing” in Suicide Note*. Nature. 13 Aug. 2014.

64 Bik E. M. et al. *The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications*. MBio. 7, no. 3 (2016): e00809–16. О самой Бик см. Bartlett T. *Hunting for Fraud Full Time*. Chronicle of Higher Education. 8 Dec. 2019.

65 Bik E. M. et al. *The Prevalence of Inappropriate Image Duplication.*

66 Например, известно, что если просить людей выбрать наугад число от 1 до 10, то они гораздо чаще остальных чисел будут выбирать се-

мерку. Если в наборе данных наблюдается непропорциональное количество семерок, это явный намек на то, что к созданию этих данных приложил руку человек. См., например, тут: *www.reddit.com/r/dataisbeautiful/comments/acow6y/asking_over_8500_students_to_pick_a_random_number.*

67 Важно отметить, что статистики используют слово “ошибка” не в отрицательном смысле — оно просто означает разницу между измеренным значением и истинным значением измеряемой величины.

68 И действительно, весь смысл многих статистических тестов состоит в том, чтобы отделить реальные эффекты — скажем, нового лекарства, испытываемого учеными, — от этой случайной ошибки выборки.

69 Haldane J. B. S. *The Faking of Genetical Results*. Eureka. 27 (1964): 21–4. Цитируется по: Pandit J. J. *On Statistical Methods to Test If Sampling in Trials Is Genuinely Random: Editorial*. Anaesthesia. 67, no. 5 (2012): 456–62.

70 Sanna L. J. et al. *Rising up to Higher Virtues: Experiencing Elevated Physical Height Uplifts Prosocial Actions*. Journal of Experimental Social Psychology. 47, no. 2 (2011): 472–6; Smeesters D., Liu J. *The Effect of Color (Red versus Blue) on Assimilation versus Contrast in Prime-to-Behavior Effects*. Journal of Experimental Social Psychology. 47, no. 3 (2011): 653–6.

71 Simonsohn U. *Just Post It: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone*. Psychological Science. 24, no. 10 (2013): 1875–88.

72 Yong E. *Uncertainty Shrouds Psychologist’s Resignation*. Nature. 12 July 2012. Также см. Seegers J. *Ontslag Hoogleraar Erasmus Na Plegen Wetenschapsfraude*. NRC Handelsblad. 25 June 2012.

73 И подобно автоматическим банковским системам обнаружения мошенничества, разрабатываются автоматические алгоритмы проверки данных, чтобы оценивать, нет ли в статьях сомнительных данных (см. пятую главу).

74 Например, есть так называемый закон Бенфорда (Benford F. *The Law of Anomalous Numbers*. Proceedings of the American Philosophical Society. 78, no. 4 (1937): 551–72; хотя эта закономерность была впервые подмечена в 1881 году математиком Саймоном Ньюкомом), математическое явление, наблюдаемое во многих разных совокупностях чисел. Закон Бенфорда гласит, что во множестве наборов данных первая значащая цифра в числах гораздо чаще будет маленькой, а не большой: с вероятностью около 30 % там будет стоять единица, с вероятностью 18 % — двойка, 13 % — тройка и так далее до девятки, которая выступает в роли первой значащей цифры только в 5 % случаев. Это справедливо для столь различных наборов данных, как количество жителей разных стран и регио-

нов, цены на недвижимость и акции, площади поверхности рек, числа из последовательности Фибоначчи. Даже количества цитирований научных статей, посвященных закону Бенфорда, распределены в соответствии с законом Бенфорда (Mir T. A. *Citations to Articles Citing Benford's Law: A Benford Analysis*. ArXiv (19 Mar. 2016): 1602.01205). Если этот закон кажется вам довольно странным и контринтуитивным, вы не одиноки: математики так и не объяснили убедительно, почему он справедлив. Тем не менее на практике он твердо установлен, и наборы данных, не подчиняющихся этому закону, хотя вообще-то должны были бы, кем-то, возможно, подделаны. Правда, идут жаркие споры о том, насколько закон Бенфорда надежен в качестве индикатора жульничества (Diekmann A., Jann B. *Benford's Law and Fraud Detection: Facts and Legends*. German Economic Review. 11, no. 3 (2010): 397–401), так что полагаться на него можно лишь как на часть полноценной совокупности методик по обнаружению мошенничества.

75 Полезная схема приводится в этой статье: Van den Bor R. M. et al. *A Computationally Simple Central Monitoring Procedure, Effectively Applied to Empirical Trial Data with Known Fraud*. Journal of Clinical Epidemiology. 87 (2017): 59–69.

76 LaCour M. J., Green D. P. *When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality*. Science. 346, no. 6215 (2014): 1366–9.

77 McGee H. *Personal Route to Reach Public Central to Yes Campaign*. Irish Times. 14 May 2015.

78 Цитируется по: Munger M. C. *L'Affaire LaCour: What It Can Teach Us about Academic Integrity and "Truthiness"*. Chronicle of Higher Education. 15 June 2015.

79 Broockman D. et al. *Irregularities in LaCour (2014)*. 19 May 2015.

80 Bartlett T. *The Unraveling of Michael LaCour*. Chronicle of Higher Education. 2 June 2015. Следует отметить, что Лакур предоставил ответ на обвинения (крайне слабый, на мой взгляд): Malakoff D. *Gay Marriage Study Author LaCour Issues Defense, but Critics Aren't Budging*. Science. 30 May 2015. В итоге Брукман и Калла опубликовали результаты своего собственного, настоящего опросного исследования, в ходе которого они проверяли некоторые гипотезы из фальшивой работы Лакура, только речь шла о правах трансгендеров, а не гомосексуалов. Они пришли к выводу, что опрашивание лицом к лицу действительно снижает предвзятость, однако неважно, является ли сам опрашивающий трансгендером или нет. Broockman D., Kalla J. *Durably Reducing Transphobia: A Field Experiment on Door-to-Door Canvassing*. Science. 352, no. 6282 (2016): 220–4.

81 Brainard J. *What a Massive Database of Retracted Papers Reveals about Science Publishing's "Death Penalty"*. Science. 25 Oct. 2018.

82 *https://retractionwatch.com/retraction-watch-database-user-guide/*
83 Cho I. et al. *Retraction*. Science. 367, no. 6474 (2020): 155.
84 *https://twitter.com/francesarnold/status/1212796266494607360*
85 Мошенничество: Grieneisen M. L., Zhang M. *A Comprehensive Survey of Retracted Articles from the Scholarly Literature*. PLOS ONE. 7, no. 10 (2012): e44118. В обзоре, посвященном отозванным статьям из области психологии, приведены сходные оценки: Stricker J., Günther A. *Scientific Misconduct in Psychology: A Systematic Review of Prevalence Estimates and New Empirical Data*. Zeitschrift Für Psychologie. 227, no. 1 (2019): 53–63. Плагиат: об этом же, в общем-то, говорится и в других расследованиях причин отзыва статей, например, см. Bozzo A. et al. *Retractions in Cancer Research: A Systematic Survey*. Research Integrity and Peer Review. 2, no. 1 (2017): 5; Corbyn Z. *Misconduct Is the Main Cause of Life-Sciences Retractions*. Nature. 490, no. 7418 (2012): 21; Li G. et al. *Exploring the Characteristics, Global Distribution and Reasons for Retraction of Published Articles Involving Human Research Participants: A Literature Survey*. Journal of Multidisciplinary Healthcare. 11 (2018): 39–47. См. общий обзор: Gross C. *Scientific Misconduct*. Annual Review of Psychology. 67, no. 1 (2016): 693–711.
86 Fanelli D. *Why Growing Retractions Are (Mostly) a Good Sign*. PLOS Medicine. 10, no. 12 (2013): e1001563.
87 Преступления в обществе: Caspi A. et al. *Childhood Forecasting of a Small Segment of the Population with Large Economic Burden*. Nature Human Behaviour. 1, no. 1 (2017): 0005. Ученые: Brainard J. *What a Massive Database of Retracted Papers Reveals about Science Publishing's "Death Penalty"*. Science. 25 Oct. 2018.
88 *https://retractionwatch.com/the-retraction-watch-leaderboard.* Сейчас для того, чтобы попасть в "список лидеров", нужно иметь за плечами минимум двадцать восемь отозванных статей.
89 Kranke P. et al. *Reported Data on Granisetron and Postoperative Nausea and Vomiting by Fujii et al. Are Incredibly Nice!* Anesthesia & Analgesia. 90, no. 4 (2000): 1004–6.
90 Marcus A., Oransky I. *How the Biggest Fabricator in Science Got Caught*. Nautilus. 21 May 2015.
91 Неправдоподобные данные: Carlisle J. B. *The Analysis of 168 Randomised Controlled Trials to Test Data Integrity.* Anaesthesia. 67, no. 5 (2012): 521–37. Конец карьере: Normile D. *A New Record for Retractions? (Part 2).* Science. 2 July 2012.
92 Marcus A. *Does Anesthesiology Have a Problem? Final Version of Report Suggests Fujii Will Take Retraction Record, with 172*. Retraction Watch. 2 July 2012.
93 Fanelli D. *How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data*. PLOS ONE. 4, no. 5 (2009): e5738.

94 Там же.
95 Gross C. *Scientific Misconduct*.
96 Это Рашми Мадхури, индийский специалист по наноматериалам. На ее счету двадцать четыре отозванных статьи, преимущественно из-за дублирования изображений. McCook A. *Author under Fire Has Eight Papers Retracted, Including Seven from One Journal*. Retraction Watch. 25 April 2018.
97 Fang F. C. et al. *Males Are Overrepresented among Life Science Researchers Committing Scientific Misconduct*. MBio. 4, no. 1 (2013): e00640–12.
98 Fanelli D. et al. *Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity*. PLOS ONE. 10, no. 6 (2015): e0127556. Вполне правдоподобно, что мужчины должны быть ответственны за большее число случаев мошенничества в науке, чем женщины: согласно данным ФБР по арестам 2017 года из-за любого вида мошеннических действий, мужчины составляли 65,5 % при арестах за подделку документов и контрафакцию и 62,5 % — за мошенничество (правда, лишь 50,9 % — за растрату имущества; Информационная служба криминальной юстиции, “Преступления в США: 2017 год”, таблица 42: *https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topic-pages/tables/table-42*). Разумеется, так обстояли дела не только с мошенничеством и связанными с ним преступлениями: ни в одной из категорий преступлений по данным ФБР женщины не превосходили мужчин в процентном соотношении (ближе всего значения для растраты имущества — разница между полами составляет всего лишь около 2 %). В целом, если учесть все категории преступлений, то, согласно этим данным, в 73 % случаев арестовывали мужчин.
99 Fanelli D. et al. *Testing Hypotheses on Risk Factors for Scientific Misconduct via Matched-Control Analysis of Papers Containing Problematic Image Duplications*. Science and Engineering Ethics. 25, no. 3 (2019): 771–89.
100 Правда, это может измениться: в 2017 году некоторые суды в Китае запрашивали высшую меру наказания (не метафорически, как в выражении про отзыв статьи как “высшую меру наказания в науке”, а буквально — смертную казнь) в делах о научном мошенничестве. Основатели *Retraction Watch* объяснили, почему это плохая идея: Oransky I., Marcus A. *Chinese Courts Call for Death Penalty for Researchers Who Commit Fraud*. STAT News. 23 June 2017.
101 Wang Y. et al. *Positive Results in Randomized Controlled Trials on Acupuncture Published in Chinese Journals: A Systematic Literature Review*. The Journal of Alternative and Complementary Medicine. 20, no. 5 (2014): A129. Цитируется по: Novella S. *Scientific Fraud in China*. Science-Based Medicine. 27 Nov. 2019.

102 Liao Q.-J. et al. *Perceptions of Chinese Biomedical Researchers Towards Academic Misconduct: A Comparison Between 2015 and 2010*. Science and Engineering Ethics. 10 April 2017.
103 Stern A. M. et al. *Financial Costs and Personal Consequences of Research Misconduct Resulting in Retracted Publications*. eLife. 3 (2014): e02956. Версия про "трудности с финансированием" изложена здесь: Chevassus-au-Louis N. *Fraud in the Lab: The High Stakes of Scientific Research.* Cambridge, MA: Harvard University Press, 2019.
104 Medawar P. *The Strange Case of the Spotted Mice.*
105 Goodstein D. *On Fact and Fraud: Cautionary Tales from the Front Lines of Science*. Princeton: Princeton University Press, 2010.
106 Schön J. H. et al. *Field-Effect Modulation of the Conductance of Single Molecules*. Science. 294, no. 5549 (2001): 2138–40.
107 Стэнфордский профессор: цитируется по *World's Smallest Transistor*. Engineer. 9 Nov. 2001. Подробную историю аферы Шёна см. здесь: Reich E. S. *Plastic Fantastic: How the Biggest Fraud in Physics Shook the Scientific World.* Basingstoke, Hampshire: Palgrave Macmillan, 2009.
108 Cassuto L. *Big Trouble in the World of "Big Physics".* The Guardian. 18 Sept. 2002.
109 *Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrick Schön and Coauthors*. American Physical Society. Sept. 2002. Еще Шён попросту не вел записей по огромной части своей лабораторной работы. См. Reich E. S. *Plastic Fantastic.*
110 *Report of the Investigation Committee.*
111 Там же.
112 Позиция Шёна в "списке лидеров" *Retraction Watch*, может, и не очень высока, но у него, вероятно, самое большое количество статей, отозванных из топовых журналов вроде *Science* и *Nature*.
113 Stapel D. A. *Faking Science*.
114 О расследовании свыше 146 случаев нарушения научной этики, проведенном американской Службой по обеспечению добросовестности в научной практике, и обсуждение некоторых общих особенностей см. здесь: Kornfeld D. S. *Perspective: Research Misconduct*. Academic Medicine. 87, no. 7 (2012): 877–82.
115 Interlandi J. *An Unwelcome Discovery*. New York Times. 22 Oct. 2006.
116 Levelt Committee et al. *Flawed Science.*
117 В одном анализе биомедицинских публикаций было показано, что в течение года после отзыва статьи она получает 45 % цитирований по сравнению со схожими неотозванными статьями, а дальше это число со временем снижается. См. Furman J. L. et al. *Governing Knowledge in the Scientific Community: Exploring the Role of Retractions in Biomedicine.* Research Policy. 41, no. 2 (2012): 276–90.

118 Bornemann-Cimenti H. et al. *Perpetuation of Retracted Publications Using the Example of the Scott S. Reuben Case: Incidences, Reasons and Possible Improvements*. Science and Engineering Ethics. 22, no. 4 (2016): 1063–72.

119 Bar-Ilan J., Halevi G. *Post Retraction Citations in Context: A Case Study*. Scientometrics. 113, no. 1 (2017): 547–65. Еще более угнетающие показатели см. здесь: Neale A. V. et al. *Analysis of Citations to Biomedical Articles Affected by Scientific Misconduct*. Science and Engineering Ethics. 16, no. 2 (2010): 251–61.

120 Возможно, ученые сохраняли копию статьи на своем компьютере или обретали в бумажном виде еще до ее отзыва, а потом уже больше ничего не проверяли. См. Teixeira da Silva J. A., Bornemann-Cimenti H. *Why Do Some Retracted Papers Continue to Be Cited?* Scientometrics. 110, no. 1 (2017): 365–70; а также Teixeira da Silva J. A. et al. *Citing Retracted Papers Has a Negative Domino Effect on Science, Education, and Society*. Impact of Social Sciences. 6 Dec. 2016.

121 На счету Йоахима Больдта 100 отозванных статей [уже 186, согласно *Retraction Watch (прим. перев.)*]. Вероятно, вы заметили, что в научное мошенничество как будто бы вовлечено немало анестезиологов. Если это не чистая случайность, то единственное разумное объяснение, которое приходит мне в голову, состоит в том, что анестезиология — до сих пор, в общем-то, достаточно таинственная и плохо понимаемая область. Это оставляет большое пространство для маневра разным мошенникам, совершающим "открытия", которые в этой области, где нет суперстрогих данных, сравнить особенно не с чем. Заметьте: это всего лишь мое предположение.

122 Zarychanski R. et al. *Association of Hydroxyethyl Starch Administration with Mortality and Acute Kidney Injury in Critically Ill Patients Requiring Volume Resuscitation: A Systematic Review and Meta-Analysis*. JAMA. 309, no. 7 (2013): 678–88.

123 И однако же — даже после отзыва ста статей Больдта — на момент написания этих строк чуть ли не сто других его публикаций остаются в научной литературе [ситуация с тех пор улучшилась: согласно информации на сайте *Retraction Watch,* отозвано уже 186 его статей *(прим. перев.)*]. Редакторы журналов находятся в незавидном положении: для многих из этих статей фальсификации напрямую не доказаны, но, учитывая, какой Больдт плодовитый обманщик, все мы понимаем, что они там наверняка есть. Было высказано одно предложение: редакторам следует добавлять к статьям подловленных мошенников примечание ("выражение озабоченности" на редакционном жаргоне) о том, что ссылаться на эти публикации нужно с превеликой осторожностью. См. Wiedermann C. J. *Inaction over Retractions of Identified Fraudulent Publications: Ongo-*

ing Weakness in the System of Scientific Self-Correction. Accountability in Research. 25, no. 4 (2018): 239–53; а также Wiedermann C. J., Joannidis M. *The Boldt Scandal Still in Need of Action: The Example of Colloids 10 Years after Initial Suspicion of Fraud*. Intensive Care Medicine. 44, no. 10 (2018): 1735–7.

124 Wakefield A. J. et al. *Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, and Pervasive Developmental Disorder in Children*. Lancet. 351, no. 9103 (1998): 637–41.

125 Wakefield A. J. et al. *Enterocolitis in Children with Developmental Disorders*. The American Journal of Gastroenterology. 95, no. 9 (2000): 2285–95. Действительно, есть свидетельства в пользу того, что у детей с расстройствами аутистического спектра чаще проявляются некоторые желудочно-кишечные симптомы (McElhanon B. O. et al. *Gastrointestinal Symptoms in Autism Spectrum Disorder: A Meta-Analysis*. Pediatrics. 133, no. 5 (2014): 872–83), однако нет никаких доказательств, что вызывается это вакцинами.

126 *Brian Deer's 2004 film on Andrew Wakefield*. Twenty Twenty Television, 2004. *https://youtu.be/7UbL8opM6TM*

127 Taylor L. E. et al. *Vaccines Are Not Associated with Autism: An Evidence-Based Meta-Analysis of Case-Control and Cohort Studies*. Vaccine. 32, no. 29 (2014): 3623–9; Golding J. et al. *Prenatal Mercury Exposure and Features of Autism: A Prospective Population Study*. Molecular Autism. 9, no. 1 (2018): 30; Dudley M. Z. et al. *The Clinician's Vaccine Safety Resource Guide: Optimizing Prevention of Vaccine-Preventable Diseases Across the Lifespan.* Cham: Springer International Publishing, 2018; Hviid A. et al. *Measles, Mumps, Rubella Vaccination and Autism: A Nationwide Cohort Study*. Annals of Internal Medicine. 170, no. 8 (2019): 513–20.

128 Dudley M. Z. et al. *The Clinician's Vaccine Safety Resource Guide*.

129 Godlee F. et al. *Wakefield's Article Linking MMR Vaccine and Autism Was Fraudulent*. BMJ. 342 (2011): c7452.

130 Deer B. *How the Case against the MMR Vaccine Was Fixed*. BMJ. 342 (2011): c5347.

131 Как подтвердит вам любой философ-первокурсник, аргумент “вакцина — затем симптомы” вообще с большой натяжкой можно было считать решающим: просто из того, что событие Y происходит вскоре после X, нельзя сделать непреложный вывод, будто X вызывает Y.

132 Deer B. *How the Vaccine Crisis Was Meant to Make Money*. BMJ. 342 (2011): c5258.

133 Дир сообщил, что Уэйкфилд получал от того адвоката 150 фунтов в час и в общей сложности заработал 435 643 фунта (не считая издержек), а были это деньги налогоплательщиков — из британского фонда *Legal Aid. https://briandeer.com/wakefield/legal-aid.htm*

134 *http://briandeer.com/wakefield/vaccine-patent.htm*
135 Wakefield A. J. et al. *Ileal-Lymphoid-Nodular Hyperplasia.*
136 Deer B. *The Lancet's Two Days to Bury Bad News*. BMJ. 342 (2011): c7001.
137 Отчет слушаний Главного медицинского совета по вопросу о пригодности к практике Уэйкфилда и двух его коллег от 16 июля 2007 года см. тут: *www.channel4.com/news/media/2010/01/day28/GMC_Charge_sheet.pdf.*
138 *Ruling on Doctor in MMR Scare*. NHS News. 29 Jan. 2010.
139 *Vaxxed: Tribeca Festival Withdraws MMR Film*. BBC News. 27 March 2016.
140 *Daily Mail* как газета-зачинщица: Goldacre B. *The MMR Sceptic Who Just Doesn't Understand Science*. Bad Science. 2 Nov. 2005; журнал *Private Eye*: Elliman D., Bedford H. *Private Eye Special Report on MMR*. BMJ. 324, no. 7347 (2002): 1224.
141 В данном случае есть серьезные основания возлагать особую вину на средства массовой информации. В исследовании 2019 года было показано, что всплеск дезинформации о вакцине MMR в медиа напрямую связан с падением показателей вакцинации (а при определенных допущениях — это падение и вызвал). Разумеется, первопричиной послужила публикация статьи Уэйкфилда. См. Tangvatcharapong M. *The Impact of Fake News: Evidence from the Anti-Vaccination Movement in the US*. Oct. 2019.
142 80 %: *Childhood Vaccination Coverage Statistics: England 2017–18*. NHS Digital. 18 Sept. 2018. Скачок показателей вверх: *Vaccine Knowledge Project. Measles.* University of Oxford. 25 June 2019.
143 Всемирная организация здравоохранения: *More than 140,000 Die from Measles as Cases Surge Worldwide*. World Health Organization. 5 Dec. 2019. Всплеск заболеваемости корью: Boseley S. *Resurgence of Deadly Measles Blamed on Low MMR Vaccination Rates*. The Guardian. 21 Aug. 2018.
144 Такой аргумент приводился некоторыми журналистами, в тот момент пребывающими на стороне страшащихся вакцины MMR, в том числе Питером Хитченсом из *Mail on Sunday* (Hitchens P. *Some Reflections on Measles and the MMR*. Peter Hitchens's Blog. 11 April 2013). Более того, во многих текстах про антивакцинаторство звучала мысль, что вакцины — жертвы собственного успеха. Корь, паротит и краснуха — три крайне неприятных болезни, и в редких случаях они дают серьезные, меняющие жизнь осложнения вроде глухоты. А вакцины столь эффективно изничтожили эти заболевания, что большинство людей в излишней самонадеянности позабыли, как страшны могут быть эти хвори.
145 Chaplin S. et al. *Wellcome Trust Global Monitor 2018*. Wellcome Trust. 19 June 2019.

146 Другая подобная история — недавний скандал с фирмой *Theranos*. Руководительница компании Элизабет Холмс — сейчас она под следствием за мошенничество [в 2022-м ее приговорили к одиннадцати годам и трем месяцам тюрьмы *(прим. перев.)*] — ухитрилась вытянуть невероятные суммы денег из инвесторов, в том числе Руперта Мёрдока, семьи Уолтонов (основателей сети магазинов *Walmart*) и многих других, и стала самой молодой и самой богатой в Америке миллиардершей, добившейся успеха собственными силами. Выпускавшиеся ее компанией приборы, якобы умеющие диагностировать множество заболеваний по капельке крови, на самом деле никогда не работали. Однако инвесторы, желавшие вложиться в дело на этапе зарождения чего-то вроде нового *Facebook* или *Uber* в смысле революционного технологического эффекта, умудрились не заметить или проигнорировать очевидные изъяны. См. увлекательнейшую книгу от журналиста-расследователя: Carreyrou J. *Bad Blood: Secrets and Lies in a Silicon Valley Startup*. New York: Alfred A. Knopf, 2018. [Каррейру Дж. *Дурная кровь. Тайны и ложь одного стартапа Кремниевой долины*. М.: АСТ, 2020.]

147 О похожей истории можно прочитать здесь: McCook A. *Two Researchers Challenged a Scientific Study About Violent Video Games — and Took a Hit for Being Right.* Vice. 25 July 2018.

148 Дальнейшее обсуждение этого вопроса см. здесь: Hilgard J. *Are Frauds Incompetent?* Crystal Prison Zone. 1 Feb. 2020.

Глава 4. **Предвзятость**

1 Schopenhauer A. *The World as Will and Presentation*. Vol. II. New York: Routledge, 2011.

2 Huxley T. H. *The Darwin Memorial* (1885) // *Collected Essays.* Cambridge: Cambridge University Press, 2011.

3 См., например: Morton S. G. *Crania Americana.* London: Simkin, Marshall & Co., 1839.

4 Или, возможно, горошинки черного перца. См. Mitchell P. W. *The Fault in His Seeds: Lost Notes to the Case of Bias in Samuel George Morton's Cranial Race Science*. PLOS Biology. 16, no. 10 (2018): e2007008.

5 Morton S. G. *Aug. 8th, 1848, Vice President Morton in the Chair*. Proceedings of the Academy of Natural Sciences of Philadelphia. 4 (1848): 75–6.

6 Gould S. J. *The Mismeasure of Man*. Rev. and Expanded. New York: Norton, 1996.

7 Gould S. J. *Morton's Ranking of Races by Cranial Capacity. Unconscious Manipulation of Data May Be a Scientific Norm*. Science. 200, no. 4341 (1978): 503–9.

8 Там же.

9 В приложении к этой статье приведена полезная классификация подобных предубеждений: Sackett D. L. *Bias in Analytic Research. The Case-Control Study Consensus and Controversy*. Journal of Chronic Diseases. 32, no. 1–2 (1979): 51–63.

10 Вообще термином “предубеждение” обозначается все то, что систематически смещает результаты прочь от истины. “Систематичность” важна: в отличие от случайных ошибок (измерения и выборки), которые мы обсуждали в предыдущей главе, у предубеждений есть направленность. Случайные ошибки сродни автомобилю с неисправным рулем, беспорядочно вихляющему то вправо, то влево. А вот предубеждение — это автомобиль с перекошенной осью, которая все время утягивает его на одну какую-то сторону. Некоторые предубеждения могут вызываться факторами, не связанными с человеком: например, сбоящими приборами или “глюками” в программном обеспечении компьютера. Но нам сейчас интересны именно те предубеждения, в коих виноваты сами ученые.

11 Fanelli D. *“Positive” Results Increase Down the Hierarchy of the Sciences*. PLOS ONE. 5, no. 4 (2010): e10068.

12 Есть некоторое расхождение во мнениях относительно того, уменьшается или увеличивается с течением времени количество положительных и отрицательных результатов. Одну точку зрения на этот вопрос см. здесь: Fanelli D. *Negative Results Are Disappearing from Most Disciplines and Countries.* Scientometrics. 90, no. 3 (2011): 891–904; другую — тут: De Winder J. C. F., Dodou D. *A Surge of p-Values between 0.041 and 0.049 in Recent Decades (but Negative Results Are Increasing Rapidly Too)*. PeerJ. 3 (2015): e733.

13 Есть и другая причина, почему уровень успеха выше 90 % служил бы плохим знаком, даже если бы он был точен и не происходило ничего подозрительного: это бы означало, что ученые настолько хорошо выбирают правильные гипотезы, что еще до того, как приступить к проверке, знают, что верно, а что нет. В подобном мире почти стопроцентного успеха ученые воздерживались бы от изучения подлинно новых, острых вопросов, ответы на которые куда более неопределенные, а исследования которых рискованнее. И таким образом они бы пренебрегали важной ролью науки — исследовать неизведанное и углублять наши знания о мире.

14 Rosenthal R. *The File Drawer Problem and Tolerance for Null Results*. Psychological Bulletin. 86, no. 3 (1979): 638–41.

15 Поскольку рост в разных странах отличается, австрийские женщины, как выясняется, в среднем выше перуанских мужчин (хотя разница между полами внутри каждой из этих стран сохраняется:

перуанки ниже своих мужчин-соотечественников, а австрийцы выше австриек). *https://en.wikipedia.org/wiki/Average_human_height_by_country#Table_of_Heights*

16 Это было бы недооценкой реального показателя: согласно *Wikipedia*, средняя разница в росте между мужчинами и женщинами в Шотландии в 2008 году составляла 13,7 сантиметра. См. ссылку из предыдущего примечания.

17 Конкретные детали того, как вычисляется *p*-значение, не обязательно нужны для понимания, как оно работает. Четкое введение в статистику как таковую см. здесь: Spiegelhalter D. *The Art of Statistics: Learning from Data*. London: Penguin, 2019. Доступное обсуждение более философских вопросов, касающихся статистики, см. тут: Dienes Z. *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. New York: Palgrave Macmillan, 2008.

18 Cassidy S. A. et al. *Failing Grade: 89% of Introduction-to-Psychology Textbooks That Define or Explain Statistical Significance Do So Incorrectly*. Advances in Methods and Practices in Psychological Science. 2, no. 3 (2019): 233–9. См. также Hubbard R., Bayarri M. J. *Confusion Over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing*. American Statistician. 57, no. 3 (2003): 171–8.

19 Позиция Американской статистической ассоциации по вопросу *p*-значений удивительно доходчиво изложена здесь: Wasserstein R. L., Lazar N. A. *The ASA Statement on p-Values: Context, Process, and Purpose*. The American Statistician. 70, no. 2 (2016): 129–33. Там дается такое определение *p*-значению: “в рамках конкретной статистической модели это вероятность того, что статистический вывод по данным (то есть различие в средних по выборке между двумя сравниваемыми группами) будет таким же или еще значительнее, чем полученный”.

20 Здесь нужно оговорить еще два момента. Во-первых, другое распространенное заблуждение относительно *p*-значения заключается в том, что оно якобы показывает, насколько вероятна “нулевая гипотеза”, то есть насколько вероятно, что в ваших данных нет никакого эффекта (и обычно вы хотите, чтобы эта вероятность была низкой). Но это не так — на самом деле, как я указал в своем определении, *p*-значение предполагает, что эффекта нет (то есть оно предполагает, что “нулевая гипотеза” верна). Оно также предполагает, что верно и много чего другого о ваших данных: есть некоторые чисто математические допущения о формах распределения переменных, а также более сложные предположения, касающиеся того, стопроцентно ли ученые честны при описании исследования и сколько всего *p*-значений они вычислили. Как мы узнаем дальше из этой главы, такие предположения часто неверны. Чем сильнее

нарушаются эти предположения, тем менее надежным и вразумительным становится *p*-значение. Во-вторых, почему в определении *p*-значения ("если на самом деле эффект нулевой, каковы были шансы все равно получить те же результаты, что у вас на руках, или демонстрирующие, казалось бы, даже еще больший эффект?") есть оговорка "еще больший эффект"? (В определении от Американской статистической ассоциации ту же роль выполняют слова "или еще значительнее".) Она необходима, поскольку вероятность получить любое конкретное значение крайне мала: представьте, например, насколько редко мы получали бы для своей выборки разницу в росте, в точности равную, скажем, 10,00144983823 сантиметра, если бы нам пришлось повторять свое исследование с шотландскими женщинами и мужчинами бесконечное число раз. Это четко определенное число было бы крайне маловероятным, независимо от того, есть ли в действительности разница в росте на уровне популяции, так что *p*-значение, которое просто говорило бы нам, насколько малы шансы получить вот это конкретное число, едва ли сильно бы нам помогало. Вот для чего нужна добавка "еще больший эффект". В нашем вымышленном примере — единожды проведенном исследовании, в котором между выбранными наугад десятью мужчинами и десятью женщинами обнаружилась разница в росте десять сантиметров, — *p*-значение, равное 0,03, означало бы, что если бы "реального" эффекта в популяции шотландцев не существовало, то вероятность обнаружить разницу в десять сантиметров или больше была бы 3 %.

21 Сразу же хочется сказать, что неплохо было бы иметь нулевые или очень близкие к нулю шансы получить ложноположительный результат. Но тут правит компромисс. Если мы будем чересчур осторожничать насчет того, какие результаты принимать, то увеличим вероятность пропустить реальные эффекты в своих наборах данных (то есть совершить ложноотрицательную ошибку).

22 Salsburg D. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*. New York: Holt, 2002.

23 *www.taps-aff.co.uk.* Создатель этого чудесного сервиса, который на самом деле учитывает не просто температуру, а куда больше показателей, — Колин Уодделл.

24 То же было предложено в статье, которая стала частью серьезной дискуссии об уровне значимости. Lakens D. et al. *Justify Your Alpha*. Nature Human Behaviour. 2, no. 3 (2018): 168–71.

25 Spiegelhalter D. *Explaining 5-Sigma for the Higgs: How Well Did They Do?* Understanding Uncertainty. 8 July 2012.

26 Dawkins R. *The Tyranny of the Discontinuous Mind*. New Statesman. 19 Dec. 2011. Обсуждение этого вопроса также можно найти в прекрасной книге Докинза: Dawkins R., Wong Y. *The Ancestor's*

Tale: A Pilgrimage to the Dawn of Life. London: Weidenfeld & Nicolson, 2016. [Докинз Р. *Рассказ предка. Паломничество к истокам жизни*. М.: Corpus, 2015.]

27 Влиянию вакцинации на снижение смертности от болезни был посвящен самый первый в истории метаанализ по медицинской теме — его провел статистик Карл Пирсон в 1904 году (для брюшного тифа), хотя методика еще не получила названия “метаанализ”. Pearson K. *Report on Certain Enteric Fever Inoculation Statistics*. BMJ. 2, no. 2288 (1904): 1243–6. Познавательная история и обзор метаанализа изложены здесь: Gurevitch J. et al. *Meta-Analysis and the Science of Research Synthesis*. Nature. 555, no. 7695 (2018): 175–82. Изменение климата: Challinor A. J. et al. *A Meta-Analysis of Crop Yield under Climate Change and Adaptation*. Nature Climate Change. 4, no. 4 (2014): 287–91.

28 Так будет, если все остальные условия совершенно одинаковы. Однако на точность результатов исследования помимо размера выборки влияет еще много других факторов, например качество измерений. Хотя размер выборки обычно служит неплохим показателем, сейчас большинство специалистов по метаанализу используют более непосредственную меру точности результата — так называемое стандартное отклонение (стандартную ошибку среднего). Именно его чаще всего откладывают по вертикальной оси на воронкообразных диаграммах.

29 При вычислении *p*-значений более существенные флуктуации небольших исследований учитываются. Если бы в действительности между мужчинами и женщинами в Шотландии различия в росте не было, то маловероятно было бы обнаружить в выборке столь значительную разницу, как десять сантиметров, однако это все равно могло бы получиться случайно при измерениях в маленькой выборке. Напомню, что *p*-значение для разницы в десять сантиметров в нашей выборке из десяти мужчин и десяти женщин равнялось 0,03. Вероятность обнаружить такую же разницу в десять сантиметров в выборке из тысячи мужчин и тысячи женщин была бы чрезвычайно мала, так что *p*-значение было бы мизерным (скажем, 0,0000001 или меньше). В таком случае у нас имелось бы более надежное свидетельство того, что в популяции действительно есть “истинный” эффект. Этот пример показывает нам, почему *p*-значение не является мерой величины или важности результата: одинаковая величина эффекта может давать разные *p*-значения — в зависимости от размера выборки.

30 Shanks D. R. et al. *Romance, Risk, and Replication: Can Consumer Choices and Risk-Taking Be Primed by Mating Motives?* Journal of Experimental Psychology: General. 144, no. 6 (2015): e142–58. Другой пример, на сей раз касающийся денежного прайминга, со сход-

ными результатами: Lodder P. et al. *A Comprehensive Meta-Analysis of Money Priming*. Journal of Experimental Psychology: General. 148, no. 4 (2019): 688–712.

31 Kyzas P. A. et al. *Almost All Articles on Cancer Prognostic Markers Report Statistically Significant Results*. European Journal of Cancer. 43, no. 17 (2007): 2559–79.

32 Tzoulaki I. et al. *Bias in Associations of Emerging Biomarkers with Cardiovascular Disease*. JAMA Internal Medicine. 173, no. 8 (2013): 664.

33 См. Turner E. H. et al. *Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy*. New England Journal of Medicine. 358, no. 3 (2008): 252–60. На момент написания этих строк самый последний метаанализ, посвященный антидепрессантам, действительно показал (умеренное) влияние на симптомы депрессии: Cipriani A. et al. *Comparative Efficacy and Acceptability of 21 Antidepressant Drugs for the Acute Treatment of Adults with Major Depressive Disorder: A Systematic Review and Network Meta-Analysis*. Lancet. 391, no. 10128 (2018): 1357–66.

34 Onishi A., Furukawa T. A. *Publication Bias Is Underreported in Systematic Reviews Published in High-Impact-Factor Journals: Metaepidemiologic Study*. Journal of Clinical Epidemiology. 67, no. 12 (2014): 1320–6.

35 Herrmann D. et al. *Statistical Controversies in Clinical Research: Publication Bias Evaluations Are Not Routinely Conducted in Clinical Oncology Systematic Reviews*. Annals of Oncology. 28, no. 5 (2017): 931–7.

36 Есть целый ряд методов, которые позволяют внести поправку к величине эффекта, если в своем метаанализе вы обнаружили признаки публикационного смещения. Я всегда применяю их с легкой тревогой, поскольку по сути это предположения (о том, насколько вам нужно уменьшить величину эффекта), основанные на предположениях (насколько публикационное смещение повлияло тут на результат). Подробнее см., например: Carter E. C. et al. *Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods*. Advances in Methods and Practices in Psychological Science. 2, no. 2 (2019): 115–44.

37 Cressey D. *Tool for Detecting Publication Bias Goes under Spotlight*. Nature. 31 March 2017; Morey R. *Asymmetric Funnel Plots without Publication Bias*. BayesFactor. 9 Jan. 2016.

38 Franco A. et al. *Publication Bias in the Social Sciences: Unlocking the File Drawer*. Science. 345, no. 6203 (2014): 1502–5.

39 *www.tessexperiments.org.* Бонус этой программы в том, что все заявленные исследования проходят рецензирование с учетом специальных критериев, поэтому те, что в итоге отбираются и воплощаются, все высокого качества и характеризуются высокой статистической

мощностью (обсуждение статистической мощности и ее значения см. в пятой главе).

40 Эти значения вычислены по таблице 2 из статьи Franco A. et al. *Publication Bias in the Social Sciences*: числа из столбца “Опубликованы” были поделены на общее количество статей из нижней строки.

41 Все цитаты приводятся по дополнительным материалам (таблице S6) к статье Franco A. et al. *Publication Bias in the Social Sciences.*

42 Вывод, сделанный в статье Франко о публикационном смещении, находит подтверждение здесь: Dwan K. et al. *Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias*. PLOS ONE. 3, no. 8 (2008): e3081. Другая классическая статья на эту тему: Chan A.-W. et al. *Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles*. JAMA. 291, no. 20 (2004): 2457–65.

43 Churchill W. *The World Crisis*. Vol. III. Part 1. London: Penguin, 2007. [Черчилль У. С. *Мировой кризис* // В 6-ти т. Т. 3. 1916–1918 гг. М.: Принципиум, 2015.]

44 Продвижение Буша: Lang S. S. *Wansink Accepts 14-Month Appointment as Executive Director of USDA Center for Nutrition Policy and Promotion*. Cornell Chronicle. 20 Nov. 2007. “Обеды с умом”: *www.investors.com/politics/editorials/smarter-lunchroom-junk-science.*

45 *The 2007 Ig Nobel Prize Winners*. 4 Oct. 2007; *www.improbable.com/ig/winners/#ig2007.* Статья про суповую миску: Wansink B., Cheney M. M. *Super Bowls: Serving Bowl Size and Food Consumption*. JAMA. 293, no. 14 (2005): 1727–8. Про эту статью рассказывалось в известной книге Ричарда Талера и Касса Санстейна *Nudge* 2008 года, описана она была так: “очередной шедевр Вонсинка”. А с тех пор Талер получил настоящую Нобелевскую премию по экономике (это ее часто используемое неофициальное название, а для педантов приведу и полное: “Премия Шведского национального банка по экономическим наукам памяти Альфреда Нобеля”). Thaler R. H., Sunstein C. R. *Nudge: Improving Decisions about Health, Wealth and Happiness.* New Haven: Yale University Press, 2008. [Талер Р., Санстейн К. *Nudge. Архитектура выбора. Как улучшить наши решения о здоровье, благосостоянии и счастье.* М.: Манн, Иванов и Фербер, 2017.]

46 Исследование на тему размера порций: Wansink B., Cheney M. M. *Super Bowls*. Закупка продуктов на голодный желудок: Tal A., Wansink B. *Fattening Fasting: Hungry Grocery Shoppers Buy More Calories, Not More Food*. JAMA Internal Medicine. 173, no. 12 (2013): 1146–8. Персонажи на упаковках с хлопьями: Musicus A. et al. *Eyes in the Aisles: Why Is Cap’n Crunch Looking Down*

at My Child? Environment and Behavior. 47, no. 7 (2015): 715–33. Еще Вонсинк снял серию видео для продвижения своей работы, некоторые доступны на *YouTube*. Например, вот здесь объясняется исследование про хлопья: *www.youtube.com/watch?v=8u6xdGCIq6o.* А вот тут разъясняется очень подробно — пожалуй, гораздо подробнее, чем оно того заслуживает, — почему исследование про хлопья смехотворно: Simaneck D.E. *Debunking a Shoddy "Research" Study*. Donald Simanek's Skeptical Documents and Links. April 2014. Элмо на яблоках: Wansink B. et al. *Can Branding Improve School Lunches?* Archives of Pediatrics & Adolescent Medicine. 166, no. 10 (2012): 967–8.

47 Этот пост уже удален, но интернет ничего не забывает, так что с помощью архива *Wayback Machine* (*http://web.archive.org/web/20170312041524)* то сообщение все еще можно прочитать: *www.brianwansink.com/phd-advice/the-grad-student-who-never-said-no*

48 Aschwanden C. *We're All "P-Hacking" Now*. Wired. 26 Nov. 2019.

49 Simmons J. P. et al. *False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant*. Psychological Science. 22, no. 11 (2011): 1359–66.

50 Kerr N. L. *HARKing: Hypothesizing After the Results Are Known*. Personality and Social Psychology Review. 2, no. 3 (1998): 196–217.

51 Происхождение идеи про "техасского стрелка" обсуждается, например, здесь: Popik B. *Texas Sharpshooter Fallacy*. The Big Apple. 9 March 2013.

52 Если вероятность совершить ложноположительную ошибку равна 0,05, то вероятность *избежать* этой ошибки (то есть справедливо заявить, что эффекта нет): 1 минус 0,05. Вероятность не совершить эту ошибку в серии из *n* испытаний равняется предыдущему значению в степени *n*: $(1-0{,}05)^n$. Отсюда следует, что вероятность совершить хотя бы одну ложноположительную ошибку в серии из *n* испытаний: $1-(1-0{,}05)^n$. Поэтому если мы проведем 5 испытаний, то получим вероятность $1-(1-0{,}05)^5 = 0{,}226$, или 22,6 %. Технически это справедливо только для *независимых* испытаний — ситуаций, когда переменные, вовлеченные в каждое из них, абсолютно друг с другом не связаны. На практике же (и особенно во многих случаях *p*-хакинга, когда одни и те же переменные используются снова и снова) увеличение вероятности совершить хотя бы одну ложноположительную ошибку как функции, зависящей от числа испытаний, не будет столь же резким — однако вероятность тоже будет становиться все выше и выше, так что применим сходный принцип.

53 Еще стоит сказать, что существует целый ряд способов подправить порог для ваших *p*-значений, если вы навычисляли их много: вы можете признавать "значимыми" только те, что меньше 0,01, напри-

мер, а не 0,05. Проблема в том, что большинство исследователей забывают это сделать — либо, когда они промышляют *p*-хакингом, им не кажется, что они *действительно* прогнали столько тестов, даже если так оно и было. Есть также интересный философский вопрос: для скольких *p*-значений ученый должен делать поправку? Для каждого *p*-значения, которое он вычислил в рамках этой конкретной статьи? Для каждого *p*-значения, вычисленного при изучении данной темы? Для каждого *p*-значения, посчитанного за всю его карьеру? А что насчет всех тех *p*-значений, которые еще только будут вычислены в будущем? Простого ответа нет, как и на любой другой интересный философский вопрос. Одна точка зрения представлена здесь: Lakens D. *Why You Don't Need to Adjust Your Alpha Level for All Tests You'll Do in Your Lifetime*. The 20% Statistician. 14 Feb. 2016.

54 Аналогия взята отсюда: McIntyre L. *The Scientific Attitude: Defending Science from Denial, Fraud, and Pseudoscience.* Cambridge, Massachusetts: The MIT Press, 2019.

55 Это далеко не новое утверждение, вот статья 1969 года как раз о том же: Armitage P. et al. *Repeated Significance Tests on Accumulating Data*. Journal of the Royal Statistical Society, Series A (General). 132, no. 2 (1969): 235–44.

56 Признание: один из исследователей, который разбирал статьи Вонсинка, — Ник Браун, мой коллега и друг.

57 Van der Zee T. et al. *Statistical Heartburn: An Attempt to Digest Four Pizza Publications from the Cornell Food and Brand Lab*. BMC Nutrition. 3, no. 1 (2017): 54.

58 *Notice of Retraction: The Joy of Cooking Too Much: 70 Years of Calorie Increases in Classic Recipes.* Annals of Internal Medicine. 170, no. 2 (2019): 138.

59 Wansink B. et al. *Notice of Retraction and Replacement. Wansink B, Just DR, Payne CR. Can Branding Improve School Lunches? Arch Pediatr Adolesc Med. 2012; 166 (10): 967–968.* JAMA Pediatrics. 21 Sept. 2017. Ошибки впервые были описаны здесь: Brown N. J. L. *A Different Set of Problems in an Article from the Cornell Food and Brand Lab*. Nick Brown's Blog. 15 Feb. 2017. Другой критик Вонсинка составил досье со всеми статьями, где обнаружились ошибки: Van der Zee T. *The Wansink Dossier: An Overview*. The Skeptical Scientist. 21 March 2017.

60 Помимо тех восемнадцати отозванных статей еще и многие другие были исправлены или обзавелись “выражением озабоченности”. Их можно найти по имени автора *(Brian Wansink)* в надежной базе данных *Retraction Watch*: *http://retractiondatabase.org.* Особенно досталось статье про наклейки с Элмо: после того как она была отозвана, редакторы журнала позволили Вонсинку опубликовать исправленную версию, но едва она вышла, как в ней нашли другую

серьезную ошибку (там указывалось, что детям, участвовавшим в исследовании, было от восьми до одиннадцати лет, а на самом деле их возраст был от трех до пяти). Эта статья-замена была отозвана — и тем самым стала единственной из известных мне научных публикаций, приговоренной к “высшей мере наказания в науке” дважды. Wansink B. et al. *Notice of Retraction. Wansink B, Just DR, Payne CR. Can Branding Improve School Lunches? Arch Pediatr Adolesc Med. 2012; 166 (10): 967–968.* JAMA Pediatrics. 171, no. 12 (2017): 1230.

61 В ходе расследования, проведенного Корнеллским университетом, выяснилось, что Вонсинк “нарушал академические нормы в своей научной работе и преподавательской деятельности, в том числе заявлял неверные данные исследований, использовал сомнительные статистические методы, не документировал и не хранил результаты исследований должным образом и указывал ненадлежащее авторство”. Kotlikoff M. *Provost Issues Statement on Wansink Academic Misconduct Investigation*. Cornell Chronicle. 20 Sept. 2018.

62 Lee S. M. *Here's How Cornell Scientist Brian Wansink Turned Shoddy Data into Viral Studies about How We Eat*. BuzzFeed News. 25 Feb. 2018.

63 Увы и ах, во многих случаях, делая эти свои запросы, они вообще неделикатны, как свидетельствуют исследователи из проекта *Bullied into Bad Science* (“Принуждение к некорректной науке”): *http://bulliedintobadscience.org.*

64 Carney D. R. *My Position on “Power Poses”*. 26 Sept. 2016. *http://faculty.haas.berkeley.edu/dana_carney/pdf_My%20position%20on%20power%20poses.pdf*

65 *https://twitter.com/nicebread303/status/780395235268501504; https://twitter.com/PeteEtchells/status/780425109077106692; https://twitter.com/cragcrest/status/780447545126293504; https://twitter.com/timothycbates/status/780386384276230144; https://twitter.com/MichelleNMeyer/status/780437722393698305; https://twitter.com/eblissmoreau/status/780594280377176064.*

66 Цитируется по: Singal J., Dahl M. *Here Is Amy Cuddy's Response to Critiques of Her Power-Posing Research*. The Cut. 30 Sept. 2016.

67 John L. K. et al. *Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling*. Psychological Science. 23, no. 5 (2012): 524–32. Я привожу значения, находящиеся посередине между “уровнем самостоятельного признания” в контрольной группе и в экспериментальной (см. табл. 1 в этой статье), где людям обещали, что если они скажут правду, то организаторы пожертвуют деньги на благотворительность, так что там уровень признания в нечестных методах проведения исследований оказался выше. В этом исследовании участвовали психологи из США, но похожие результаты были получены и в Италии: Agnoli F. et al. *Ques-*

tionable Research Practices among Italian Research Psychologists. PLOS ONE. 12, no. 3 (2017): e0172792. См. также опрос, проведенный в Германии, в котором оспорены некоторые формулировки вопросов из исходного американского исследования: Fiedler K., Schwarz N. *Questionable Research Practices Revisited*. Social Psychological and Personality Science. 7, no. 1 (2016): 45–52.

68 Wang M. Q. et al. *Identifying Bioethical Issues in Biostatistical Consulting: Findings from a US National Pilot Survey of Biostatisticians*. BMJ Open. 7, no. 11 (2017): e018491. Сложив соответствующие числа, я подсчитал общий процент статистиков, сообщивших о том, что им поступали такие просьбы от одного до девяти раз либо более десяти раз за последние пять лет.

69 Necker S. *Scientific Misbehavior in Economics*. Research Policy. 43, no. 10 (2014): 1747–59. Поразительно, но в этом же опросе 2 % экономистов еще признались, что “принимали или предлагали секс в обмен на (со)авторство, доступ к данным или поддержку конкретных лиц”.

70 Masicampo E. J., Lalande D. R. *A Peculiar Prevalence of p Values Just Below.05*. Quarterly Journal of Experimental Psychology. 65, no. 11 (2012): 2271–9. См. также Barnett A. G., Wren J. D. *Examination of CIs in Health and Medical Journals from 1976 to 2019: An Observational Study*. BMJ Open. 9, no. 11 (2019): e032506. Здесь тот же вопрос изучался с помощью доверительных интервалов, еще одного статистического понятия, которое по сути дает ту же информацию, что и *p*-значение. Между прочим, схожая ситуация наблюдается на диаграммах, отражающих результаты школьных экзаменов, где для значений чуть выше установленного проходного балла характерен резкий скачок: доброжелательные учителя добавили проваливающимся детям как раз столько баллов, сколько было необходимо, чтобы те экзамен сдали. См. Dubner S. J. *Another Case of Teacher Cheating, or Is It Just Altruism?* Freakonomics. 7 July 2011.

71 Silberzahn R. et al. *Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results*. Advances in Methods and Practices in Psychological Science. 1, no. 3 (2018): 337–56; Landy J. F. et al. *Crowdsourcing Hypothesis Tests: Making Transparent How Design Choices Shape Research Results*. Psychological Bulletin. 16 Jan. 2020.

72 Yarkoni T., Westfall J. *Choosing Prediction Over Explanation in Psychology: Lessons from Machine Learning*. Perspectives on Psychological Science. 12, no. 6 (2017): 1100–22.

73 Gelman A., Loken E. *The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No “Fishing Expedition” or “p-Hacking” and the Research Hypothesis Was Posited Ahead of Time*. 4 Nov. 2013; Borges J. L. *The Garden of Forking Paths // La-*

byrinths. New York: New Directions, 1964. [Борхес Х. Л. *Сад расходящихся тропок* // Собр. соч. в 3-х т. Т. 1. М.: Полярис, 1994.]

74 Эта формулировка проблемы с *p*-хакингом взята из работы Yarkoni T., Westfall J. *Choosing Prediction Over Explanation in Psychology*, где авторы называют *p*-хакинг “методической переподгонкой”.

75 Giner-Sorolla R. *Science or Art? How Aesthetic Standards Grease the Way Through the Publication Bottleneck but Undermine Science*. Perspectives on Psychological Science. 7, no. 6 (2012): 567–71.

76 O'Boyle E. H. et al. *The Chrysalis Effect: How Ugly Initial Results Metamorphosize into Beautiful Articles*. Journal of Management. 43, no. 2 (2017): 376–99. Недавно эффект “окукливания” был подтвержден для психологических исследований: Cairo A. H. et al. *Gray (Literature) Matters: Evidence of Selective Hypothesis Reporting in Social Psychological Research*. Personality and Social Psychology Bulletin. 24 Feb. 2020.

77 Один из классических примеров такого дурного совета приведен в указанной ниже книге, в главе не кого иного, как Дэрила Бема (чье исследование паранормального мне воспроизвести не удалось, о чем мы говорили в предисловии). Он просит молодых научных сотрудников “вылавливать что-нибудь — что угодно — интересное” в своих данных. Конечно, рассуждает он, иногда это может привести к ложноположительным результатам, но нам все-таки следует “ошибаться в сторону открытий”. Bem D. J. *Writing the Empirical Journal Article* // Darley J. M. et al. (eds.) *The Compleat Academic: A Career Guide*. Washington, D. C.: American Psychological Association, 2003.

78 Hossenfelder S. *Lost in Math: How Beauty Leads Physics Astray*. New York: Basic Books, 2018. [Хоссенфельдер С. *Уродливая Вселенная. Как поиски красоты заводят физиков в тупик*. М.: Бомбора, 2021.] Также Хоссенфельдер утверждает, что сложные философские и теоретические проблемы в физике “заметают под ковер”, поскольку физики предпочитают вопросы, которые “с большей вероятностью дадут пригодные для публикации результаты за короткое время”. Мы еще поговорим об этом дальше. Обсуждение проблем в физике см. здесь: Smolin L. *The Trouble with Physics: The Rise of String Theory, the Fall of a Science and What Comes Next*. London: Allen Lane, 2007; Woit P. *Not Even Wrong: The Failure of String Theory and the Continuing Challenge to Unify the Laws of Physics*. London: Vintage Books, 2007.

79 De Angelis C. et al. *Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors*. New England Journal of Medicine. 351, no. 12 (2004): 1250–1.

80 Мы подробнее обсудим предварительную регистрацию исследований в восьмой главе.

81 *http://compare-trials.org*

82 Команда Голдакра еще пыталась опубликовать в журналах заметки, в которых говорилось, что результаты клинических испытаний не были изложены точно. Большинство редакторов интереса не проявили. Goldacre B. *Make Journals Report Clinical Trials Properly*. Nature. 530, no. 7588 (2016): 7.

83 Jones P. M. et al. *Comparison of Registered and Reported Outcomes in Randomized Clinical Trials Published in Anesthesiology Journals*. Anesthesia & Analgesia. 125, no. 4 (2017): 1292–1300; см. также Altman D. G. et al. *Harms of Outcome Switching in Reports of Randomised Trials: CONSORT Perspective*. BMJ. 356 (2017): j396.

84 Книги о решении проблем с медицинскими испытаниями: Goldacre B. *Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients.* London: Fourth Estate, 2012; Harris R. F. *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions*. New York: Basic Books, 2017.

85 В метаанализе это прозвали принципом “мусор на входе — мусор на выходе”. Hunt M. *How Science Takes Stock: The Story of Meta-Analysis*. New York: Russell Sage Foundation, 1998.

86 Абсолютное число финансируемых индустрией медицинских исследований со временем увеличивалось, однако их доля среди всех клинических испытаний сокращалась. Ehrhardt S. et al. *Trends in National Institutes of Health Funding for Clinical Trials Registered in ClinicalTrials.Gov*. JAMA. 314, no. 23 (2015): 2566–7.

87 Тут можно было бы возразить, что у фармацевтических компаний денег и других ресурсов достаточно, чтобы оплачивать исследования получше качеством, поэтому мы бы ожидали разных результатов. Однако будь исследования получше качеством, это означало бы, что они меньше подвержены предубеждениям, а значит, *реже* давали бы положительные результаты, ведь мы знаем, что основные предубеждения в подобных исследованиях ведут к большему числу ложноположительных результатов. (Если средство лечения работает, беспристрастное исследование покажет только это, ни больше ни меньше, а пристрастное — как преувеличит эффект, так и потенциально даст еще и какие-то ложноположительные результаты.) Кроме того, в недавнем обзоре размер выборки учитывался — вот чем, в частности, исследования, щедро финансируемые фармкомпаниями, нередко отличаются от исследований, финансируемых из других источников, — и все равно получилось, что первые чаще давали положительные результаты. Waldorff S. *Results of Clinical Trials Sponsored by For-Profit vs Nonprofit Entities*. JAMA. 290, no. 23 (2003): 3071.

88 Lathyris D. N. et al. *Industry Sponsorship and Selection of Comparators in Randomized Clinical Trials*. European Journal of Clinical

Investigation. 40, no. 2 (2010): 172–82; Estellat C. *Lack of Head-to-Head Trials and Fair Control Arms: Randomized Controlled Trials of Biologic Treatment for Rheumatoid Arthritis*. Archives of Internal Medicine. 172, no. 3 (2012): 237–44.

89 Jones C. W. et al. *Non-Publication of Large Randomized Clinical Trials: Cross Sectional Analysis*. BMJ. 347 (2013): f6104. Впрочем, стоит отметить, что "переключение на другие исходы" в исследованиях, оплачиваемых фармкомпаниями, похоже, происходит не чаще. (Jones C. W. et al. *Primary Outcome Switching among Drug Trials with and without Principal Investigator Financial Ties to Industry: A Cross-Sectional Study*. BMJ Open. 8, no. 2 (2018): e019831.) В одном обзоре даже обнаружилось, что в исследованиях, финансируемых некоммерческими организациями, дело с "переключением на другие исходы" обстоит хуже: Delgado A. F., Delgado A. F. *Outcome Switching in Randomized Controlled Oncology Trials Reporting on Surrogate Endpoints: A Cross-Sectional Analysis*. Scientific Reports. 7, no. 1 (2017): 9206.

90 Я иногда посещаю цикл лекций в своем университете, и там предусмотрен ланч, всегда четко обозначенный как "ланч не от фармкомпаний". Вот как медицинские исследователи боятся невольно принять что-либо в дар от фармацевтических компаний — подарки, которые пришлось бы указывать в разделе "Конфликт интересов" каждой своей статьи.

91 См. Chivers T. *Does Psychology Have a Conflict-of-Interest Problem?* Nature. 571, no. 7763 (2019): 20–3.

92 См. Bero L. A., Grundy Q. *Why Having a (Nonfinancial) Interest is Not a Conflict of Interest*. PLOS Biology. 14, no. 12 (2016): e2001221. Контраргументом к моему звучат рассуждения Биро и Гранди о том, что финансовые конфликты интересов и интеллектуальные — вещи разные и уподоблять их друг другу — значит "вносить неразбериху". Как бы там ни было, в следующий раз при написании научной статьи, имеющей хоть какое-то отношение к кризису воспроизводимости, должен ли я добавить нечто в духе "Я написал о кризисе воспроизводимости книгу, и мне было бы несколько неловко, если бы оказалось, что наука на самом деле в отличной форме"? Поистине убедительный довод именно так и сделать.

93 Begley S. *The Maddening Saga of How an Alzheimer's "Cabal" Thwarted Progress toward a Cure for Decades*. STAT News. 25 June 2019.

94 Huang Y.-M. et al. *Major Clinical Trials Failed the Amyloid Hypothesis of Alzheimer's Disease*. Journal of the American Geriatrics Society. 67, no. 4 (2019): 841–4; Panza F. et al. *A Critical Appraisal of Amyloid-β-Targeting Therapies for Alzheimer Disease*. Nature Reviews Neurology. 15, no. 2 (2019): 73–88.

95 Herrup K. *The Case for Rejecting the Amyloid Cascade Hypothesis*. Nature Neuroscience. 18, no. 6 (2015): 794–9.

96 Harrison J. R., Owen M. J. *Alzheimer's disease: The amyloid hypothesis on trial*. British Journal of Psychiatry. 208, no. 1 (2016): 1–3.

97 McCartney G. et al. *Why the Scots Die Younger: Synthesizing the Evidence*. Public Health. 126, no. 6 (2012): 459–70. Обсуждение этого необычного заявления о конфликте интересов см. здесь: McCartney G. L. et al. *When Do Your Politics Become a Competing Interest?* BMJ. 342 (2011): d269.

98 Однако его политические убеждения таких чувств не вызывают. Раз уж мы заговорили откровенно, мои политические убеждения таковы: сверхвысокий социальный либерализм, умеренно высокий экономический либерализм. Если что, на сайте *www.politicalcompass.org* я обычно оказываюсь где-то с левой стороны нижнего правого квадранта.

99 Duarte J. L. et al. *Political Diversity Will Improve Social Psychological Science*. Behavioral and Brain Sciences. 38 (2015): e130.

100 В одном опросе исследователи выявили корреляцию между тем, насколько психологи были либеральны, и тем, насколько они были готовы ущемлять интересы предполагаемых консерваторов при рецензировании статей или принятии решений о приеме на работу. См. Inbar Y., Lammers J. *Political Diversity in Social and Personality Psychology*. Perspectives on Psychological Science. 7, no. 5 (2012): 496–503. Любопытно, однако, что в 2019 году различий в воспроизводимости исследований, относящихся к “либеральным” либо же “консервативным” (имеется в виду, кому больше придутся по душе сделанные выводы), обнаружено не было. Reinero D. A. et al. *Is the Political Slant of Psychology Research Related to Scientific Replicability?* Preprint, PsyArXiv (7 Feb. 2019).

101 Jussim L. *Is Stereotype Threat Overcooked, Overstated, and Oversold?* Rabble Rouser. 30 Dec. 2015.

102 Flore P. C., Wicherts J. M. *Does Stereotype Threat Influence Performance of Girls in Stereotyped Domains? A Meta-Analysis*. Journal of School Psychology. 53, no. 1 (2015): 25–44; Flore P. C. et al. *The Influence of Gender Stereotype Threat on Mathematics Test Scores of Dutch High School Students: A Registered Report*. Comprehensive Results in Social Psychology. 3, no. 2 (2018): 140–74. О дополнительных доказательствах наличия публикационного смещения в исследованиях, посвященных угрозе подтверждения стереотипа, см. здесь: Shewach O. R. et al. *Stereotype Threat Effects in Settings with Features Likely versus Unlikely in Operational Test Settings: A Meta-Analysis*. Journal of Applied Psychology. 104, no. 12 (2019): 1514–34.

103 Позже те же авторы провели предварительно зарегистрированный по всем правилам масштабный эксперимент, посвященный угрозе подтверждения стереотипа и разнице между полами в успеваемости по математике. Угроза подтверждения стереотипа ни-

как не проявилась. Flore P. C. et al. *The Influence of Gender Stereotype Threat.*

104 Moss-Racusin C. A. et al. *Gender Bias Produces Gender Gaps in STEM Engagement*. Sex Roles. 79, no. 11–12 (2018): 651–70. Много других примеров см. тут: Ceci S. J. et al. (eds.) *The Underrepresentation of Women in Science: International and Cross-Disciplinary Evidence and Debate*. Frontiers Research Topics: Frontiers Media SA, 2018.

105 Becker J. B. et al. *Female Rats Are Not More Variable than Male Rats: A Meta-Analysis of Neuroscience Studies*. Biology of Sex Differences. 7, no. 1 (2016): 34.

106 Karp N. A. et al. *Prevalence of Sexual Dimorphism in Mammalian Phenotypic Traits*. Nature Communications. 8, no. 1 (2017): 15475.

107 Shansky R. M. *Are Hormones a "Female Problem" for Animal Research?* Science. 364, no. 6443 (2019): 825–6. Как объясняет Шански, сейчас многие журналы и финансирующие организации обязывают ученых включать в свои исследования и самцов, и самок. См. также Clayton J. A. *Applying the New SABV (Sex as a Biological Variable) Policy to Research and Clinical Care*. Physiology & Behavior. 187 (2018): 2–5.

108 Fine C. *Testosterone Rex: Unmaking the Myths of Sex of Our Gendered Minds*. London: Icon Books, 2017.

109 Fine C. *Feminist Science: Who Needs It?* Lancet. 392, no. 10155 (2018): 1302–3.

110 Там же. Это вариант так называемой позиционной теории: восходящей к трудам Карла Маркса философской позиции, согласно которой собственная идентичность и опыт человека (для Маркса речь шла об идентичности и опыте рабочего класса) — вот что формирует его взгляд на реальность, поэтому нам следует внимательно прислушиваться к мнению маргинальных людей, чью точку зрения мы иначе упустим из виду. См. раздел о позиционной теории (*Standpoint Theory*) тут: Anderson E. *Feminist Epistemology and Philosophy of Science*. The Stanford Encyclopedia of Philosophy. Spring 2020.

111 Lewis J. E. et al. *The Mismeasure of Science: Stephen Jay Gould versus Samuel George Morton on Skulls and Bias*. PLOS Biology. 9, no. 6 (2011): e1001071. Новые измерения производились не с семенами или дробью, а с крошечными акриловыми шариками. К слову, это показывает, что мортоновские методы XIX века не так уж и отличались от золотого стандарта века XXI. До того черепа перемеряли еще в 1988 году — тот анализ в целом подтвердил измерения Мортона, но не его расистские выводы. Michael J. S. *A New Look at Morton's Craniological Research.* Current Anthropology. 29, no. 2 (1988): 349–54.

112 "Написать «Неверное измерение человека» я решил по причинам смешанным — как профессиональным, так и личным. Признаюсь прежде всего вот в каких сильных чувствах. Вырос я в семье, где

всегда принято было участвовать в борьбе за социальную справедливость". Gould S. J. *The Mismeasure of Man*.

113 Lewis J. E. et al. *The Mismeasure of Science.*

114 Weisberg M. *Remeasuring Man*. Evolution & Development. 16, no. 3 (2014): 166–78. Также см. Weisberg M., Paul D. B. *Morton, Gould, and Bias: A Comment on "The Mismeasure of Science"*. PLOS Biology. 14, no. 4 (2016): e1002444.

115 Mitchell P. W. *The Fault in His Seeds.*

116 Kaplan J. M. et al. *Gould on Morton, Redux: What Can the Debate Reveal about the Limits of Data?* Studies in History and Philosophy of Science Part C. 52 (2015): 22–31. Эта мысль также высказывается здесь: Graves J. L. *Great Is Their Sin: Biological Determinism in the Age of Genomics*. Annals of the American Academy of Political and Social Science. 661, no. 1 (2015): 24–50.

117 Само собой, расизм из них не в пример вредоноснее в смысле влияния на общество. Я здесь хочу лишь сказать, что все эти предубеждения искажают наше видение научных результатов.

Глава 5. **Недобросовестность**

1 Cotton C. C. *Lacon, or Many Things in Few Words*. London: Longmans et al, 1825.

2 Hirschman D. *Stylized Facts in the Social Sciences*. Sociological Science. 3 (2016): 604–26.

3 Исследование некоторое время размещалось в интернете в качестве препринта (что, как мы увидим в последней главе, нормально для экономики), но в итоге было опубликовано официально: Reinhart C. M., Rogoff K. S. *Growth in a Time of Debt*. American Economic Review. 100, no. 2 (2010): 573–8.

4 Осборн: Osborne G. *Mais Lecture — A New Economic Model*. 24 Feb. 2010; республиканцы: United States Senate Committee on the Budget. *Sessions, Ryan Issue Joint Statement on Jobs Report, Call for Senate Action on Budget*. 8 July 2011.

5 Krugman P. *How the Case for Austerity Has Crumbled*. New York Review of Books. 6 June 2013.

6 Herndon T. et al. *Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff*. Cambridge Journal of Economics. 38, no. 2 (2013): 257–79.

7 Рейнхарт и Рогофф ошибку в таблице признали, однако не согласились с критикой по многим другим моментам: Reinhart C. M., Rogoff K. S. *Reinhart-Rogoff Response to Critique*. Wall Street Journal. 16 April 2013.

8 Herndon T. et al. *Does High Public Debt Consistently Stifle Economic Growth?*

9 Stevenson B., Wolfers J. *Refereeing Reinhart-Rogoff Debate*. Bloomberg Opinion. 28 April 2013.

10 Nuijten M. B. *Statcheck — a Spellchecker for Statistics*. LSE Impact of Social Sciences. 28 Feb. 2018. Приложение *statcheck*: *http://statcheck.io.*

11 Nuijten M. B. et al. *The Prevalence of Statistical Reporting Errors in Psychology (1985–2013)*. Behavior Research Methods. 48, no. 4 (2016): 1205–26. Стоит заметить, что у алгоритма *statcheck* есть критики: Schmidt T. *Statcheck Does Not Work: All the Numbers. Reply to Nuijten et al. (2017)*. Preprint, PsyArXiv (22 Nov. 2017).

12 Brown N. J. L., Heathers J. A. J. *The* GRIM *Test: A Simple Technique Detects Numerous Anomalies in the Reporting of Results in Psychology*. Social Psychological and Personality Science. 8, no. 4 (2017): 363–9.

13 Можете проверить сами на калькуляторе или использовать приложение: *http://nickbrown.fr/*GRIM.

14 Festinger L., Carlsmith J. M. *Cognitive Consequences of Forced Compliance*. Journal of Abnormal and Social Psychology. 58, no. 2 (1959): 203–10.

15 На самом деле была и третья группа участников, которым платили по двадцать долларов. При опросе они говорили, что нашли задание скучным, равно как и те, кому не платили ничего, — предположительно потому, что уменьшили свой когнитивный диссонанс мыслями о доставшихся им денежках, а не подстройкой своих убеждений.

16 Heino M. *The Legacy of Social Psychology*. Data Punk. 13 Nov. 2016.

17 С января 2020 года, согласно *Google Scholar*, на статью сослались больше 4200 раз.

18 Carlisle J. B. *The Analysis of 168 Randomised Controlled Trials*. Также см. о Карлайле: Adam D. *How a Data Detective Exposed Suspicious Medical Trials*. Nature. 571, no. 7766 (2019): 462–4.

19 См. Kendall J. M. *Designing a Research Project: Randomised Controlled Trials and Their Principles*. Emergency Medicine Journal. 20, no. 2 (2003): 164–8.

20 Carlisle J. B. *The Analysis of 168 Randomised Controlled Trials*.

21 Carlisle J. B. *Data Fabrication and Other Reasons for Non-Random Sampling in 5087 Randomised, Controlled Trials in Anaesthetic and General Medical Journal.* Anaesthesia. 72, no. 8 (2017): 944–52. Одной из главных целей Карлайла было проверить, не хуже ли дела с подозрительно выглядящими исследованиями обстоят в анестезиологии, чем в других медицинских областях. Он пришел к выводу, что ошибки в анестезиологии так же скверны, как и везде.

22 Не все впечатлились методом Карлайла. Редакторы журнала *Anesthesiology* написали резко критическую статью, где указали на некоторые недостатки в статистике Карлайла и укорили его

за подозрения, что основной причиной провалов при рандомизации было мошенничество, а не ошибки (Kharasch E. D., Houle T. T. *Errors and Integrity in Seeking and Reporting Apparent Research Misconduct*. Anesthesiology. 127, no. 5 (2017): 733–7). Карлайл предоставил ответ, весьма убедительный, на мой взгляд (Carlisle J. B. *Seeking and Reporting Apparent Research Misconduct: Errors and Integrity — a Reply*. Anaesthesia. 73, no. 1 (2018): 126–8), однако это уже другой интересный пример, который иллюстрирует, что за самими наблюдателями тоже надо наблюдать. Как бы то ни было, в следующей главе мы увидим, что метод Карлайла действительно выявил негодность рандомизации в чрезвычайно важных клинических испытаниях, посвященных питанию, а значит, абсолютно ошибочным он быть не может.

23 Wicherts J. M. et al. *The Poor Availability of Psychological Research Data for Reanalysis*. American Psychologist. 61, no. 7 (2006): 726–8. Еще см. Savage C. J., Vickers A. J. *Empirical Study of Data Sharing by Authors Publishing in PLoS Journals.* PLOS ONE. 4, no. 9 (2009): e7078; Tenopir C. et al. *Data Sharing by Scientists: Practices and Perceptions*. PLOS ONE. 6, no. 6 (2011): e21101; Christensen G., Miguel E. *Transparency, Reproducibility, and the Credibility of Economics Research*. О том, что данные становятся все менее доступными со временем: Vines T. H. et al. *The Availability of Research Data Declines Rapidly with Article Age.* Current Biology. 24, no. 1 (2014): 94–7.

24 American Type Culture Collection Standards Development Organization Workgroup ASN-0002. *Cell Line Misidentification: The Beginning of the End*. Nature Reviews Cancer. 10, no. 6 (2010): 441–8 (см. хронологию на с. 444).

25 Рак толстой кишки: Zhang P. et al. *Retraction: Critical Role of Notch Signaling in Osteosarcoma Invasion and Metastasis*. Clinical Cancer Research. 19, no. 18 (2013): 5256–7. Свиньи: Milanesi E. et al. *Molecular Detection of Cell Line Cross-Contaminations Using Amplified Fragment Length Polymorphism DNA Fingerprinting Technology.* In Vitro Cellular & Developmental Biology — Animal. 39, no. 3–4 (2003): 124–30. Крысы: Phuchareon J. et al. *Genetic Profiling Reveals Cross-Contamination and Misidentification of 6 Adenoid Cystic Carcinoma Cell Lines: ACC2, ACC3, ACCM, ACCNS, ACCS and CAC2.* PLOS ONE. 4, no. 6 (2009): e6040.

26 American Type Culture Collection Standards Development Organization Workgroup ASN-0002. *Cell Line Misidentification*.

27 Horbach S. P. J. M., Halffman W. *The Ghosts of HeLa: How Cell Line Misidentification Contaminates the Scientific Literature*. PLOS ONE. 12, no. 10 (2017): e0186281. Получается, примерно в 0,8 % всех статей о клетках использовались неверно идентифицированные клеточные линии, а в 10 % содержится ссылка

на одно из проблемных исследований (самоцитирование здесь не учтено).
28 Huang Y. et al. *Investigation of Cross-Contamination and Misidentification of 278 Widely Used Tumor Cell Lines*. PLOS ONE. 12, no. 1 (2017): e0170384.
29 85 % контаминированных клеточных линий: Ye F. et al. *Genetic Profiling Reveals an Alarming Rate of Cross-Contamination among Human Cell Lines Used in China*. The FASEB Journal. 29, no. 10 (2015): 4268–72; также см. Bian X. et al. *A Combination of Species Identification and STR Profiling Identifies Cross-Contaminated Cells from 482 Human Tumor Cell Lines*. Scientific Reports. 7, no. 1 (2017): 9774. Просмотреть удручающий реестр всех известных неправильно идентифицированных клеточных линий — а их 529 на момент написания этих строк [уже 582 *(прим. перев.)*] — можно по этой ссылке: *https://iclac.org/databases/cross-contaminations.* Между прочим, стволовые клетки, якобы полученные с помощью методики STAP, о которых мы говорили в третьей главе при обсуждении истории Харуко Обокаты, были еще и неверно идентифицированы — происходили от другого вида мыши, а не от указанного в статье. Obokata H. et al. *Retraction Note: Bidirectional Developmental Potential in Reprogrammed Cells with Acquired Pluripotency*. Nature. 511, no. 7507 (2014): 112.
30 Horbach S. P. J. M., Halffman W. *The Ghosts of HeLa.*
31 Editorial. *Towards What Shining City, Which Hill?* Nature. 289, no. 5795 (1981): 212.
32 Korch C., Varella-Garcia M. *Tackling the Human Cell Line and Tissue Misidentification Problem Is Needed for Reproducible Biomedical Research*. Advances in Molecular Pathology. 1, no. 1 (2018): 209–28.
33 2010 год: American Type Culture Collection Standards Development Organization Workgroup ASN-0002. *Cell Line Misidentification*. 2012 год: Masters J. R. *End the Scandal of False Cell Lines.* Nature. 492, no. 7428 (2012): 186. 2015 год: *Announcement: Time to Tackle Cells' Mistaken Identity.* Nature. 520, no. 7547 (2015): 264. 2017 год: Fusenig N. E. *The Need for a Worldwide Consensus for Cell Line Authentication: Experience Implementing a Mandatory Requirement at the International Journal of Cancer*. PLOS Biology. 15, no. 4 (2017): e2001438. 2018 год: Eckers J. C. et al. *Identity Crisis — Rigor and Reproducibility in Human Cell Lines*. Radiation Research. 189, no. 6 (2018): 551–2.
34 Korch C., Varella-Garcia M. *Tackling the Human Cell Line.*
35 Еще, конечно, многие считают, что аморально вообще экспериментировать на животных. Когда нет другого выбора, кроме как провести исследование на животных, ученые придерживаются на-

бора особых принципов, чтобы сделать свою работу насколько возможно более этичной. Эти принципы известны как 3R: замена (*Replacement;* ученые пытаются обходиться в исследованиях без животных, например, привлекать людей, которые могут дать свое согласие), сокращение (*Reduction*; ученые пытаются извлечь максимум полезной информации с помощью как можно меньшего числа животных) и улучшение (*Refinement*; ученые следят, чтобы животным во время исследований было насколько возможно хорошо). Термин впервые введен здесь: Russell W. M. S., Burch R. L. *The Principles of Humane Experimental Technique*. Special ed. Potters Bar: UFAW, 1992. Гораздо больше информации об этих принципах и о том, как ученые пытаются им соответствовать, содержится на сайте британского национального центра, занимающегося внедрением подхода 3R в область исследований на животных: *www.nc^{3}rs.org.uk.*

36 Macleod M. R. et al. *Risk of Bias in Reports of In Vivo Research: A Focus for Improvement*. PLOS Biology. 13, no. 10 (2015): e1002273.

37 Hirst J. A. et al. *The Need for Randomization in Animal Trials: An Overview of Systematic Reviews*. PLOS ONE. 9, no. 6 (2014): e98856.

38 Разумеется, ослепление особенно важно в медицинских исследованиях на людях — *участники* не должны знать, какое лечение получают, чтобы их ожидания не повлияли на результаты. В случае, когда и от ученых, и от участников скрыта информация, которая могла бы бросить на результаты тень предвзятости, исследование называют “двойным слепым”. Когда объекты исследования — не люди, следить за тем, чтобы участники не знали, получают ли настоящее лечение или плацебо, конечно, не так важно (хотя иногда это все же принимается во внимание).

39 Еще Маклауд и его коллеги проверяли, сообщается ли в статьях о конфликте интересов (см. четвертую главу).

40 Macleod M. R. et al. *Evidence for the Efficacy of NXY-059 in Experimental Focal Cerebral Ischaemia Is Confounded by Study Quality*. Stroke. 39, no. 10 (2008): 2824–9.

41 Классический реальный пример, иллюстрирующий эту мысль, часто приводится в учебниках по статистике — речь идет о президентских выборах 1936 года в США. Сотрудники журнала *The Literary Digest* провели массовый опрос, в котором участвовало два миллиона человек, однако не сумели сделать выборку случайной, поскольку контактировали с участниками по телефону. В то время только у состоятельных граждан имелись дома телефоны, а кроме того, чтобы раздобыть телефонные номера, сотрудники журнала воспользовались списками членов разных клубов, где менее богатые граждане опять-таки встречались реже. Так выборка оказалась нерепрезентативной, поэтому вышла ошибка: сотрудники журнала

The Literary Digest неверно предсказали, что кандидат от республиканцев Альф Лэндон на голову разобьет Франклина Рузвельта. Рузвельт набрал 61 % голосов, а тот журнал вскоре прекратил свое существование. См. Lohr S. L., Brick J. M. *Roosevelt Predicted to Win: Revisiting the 1936 Literary Digest Poll*. Statistics, Politics and Policy. 8, no. 1 (2017): 65–84.

42 Simmons J. P. et al. *Life after P-Hacking*. SSRN. 2013.

43 Говоря о надежном обнаружении эффекта, авторы подразумевали широко используемый в научной литературе стандарт. Статистическая мощность обычно считается приемлемой, когда вероятность обнаружить с помощью статистического теста эффект, если он действительно существует (то есть получить *p*-значение меньше 0,05), равна 80 % или выше. Очевидно, что чем статистическая мощность больше, тем лучше, и при достаточно крупных выборках (или достаточно сильных эффектах) статистическая мощность может существенно превышать этот минимальный порог. При мощности в 80 % двадцатипроцентный шанс упустить эффект, если он и правда существует, — это вероятность ложноотрицательного результата.

44 Button K. S. et al. *Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience*. Nature Reviews Neuroscience. 14, no. 5 (2013): 365–76. См., в частности, табл. 2.

45 Разные области нейронауки, однако, порядочно отличаются друг от друга. Nord C. L. et al. *Power-up: A Reanalysis of "Power Failure" in Neuroscience Using Mixture Modeling*. Journal of Neuroscience. 37, no. 34 (2017): 8051–61.

46 Медицинские испытания: Lamberink H. J. et al. *Statistical Power of Clinical Trials Increased While Effect Size Remained Stable: An Empirical Analysis of 136,212 Clinical Trials between 1975 and 2014*. Journal of Clinical Epidemiology. 102 (2018): 123–8. Биомедицинские исследования: Dumas-Mallet E. et al. *Low Statistical Power in Biomedical Science: A Review of Three Human Research Domains*. Royal Society Open Science. 4, no. 2 (2017): 160254. Экономика: Ioannidis J. P. A. et al. *The Power of Bias in Economics Research*. Economic Journal. 127, no. 605 (2017): F236–65. Нейровизуализация: Cremers H. R. et al. *The Relation between Statistical Power and Inference in FMRI*. PLOS ONE. 12, no. 11 (2017): e0184923. Исследования в области сестринского дела: Gaskin C. J., Happell B. *Power, Effects, Confidence, and Significance: An Investigation of Statistical Practices in Nursing Research.* International Journal of Nursing Studies. 51, no. 5 (2014): 795–806. Поведенческая экология: Jennions M. D., Moller A. P. *A Survey of the Statistical Power of Research in Behavioral Ecology and Animal Behavior.* Behavioral Ecology. 14, no. 3 (2003): 438–45. Психология: Szucs D., Ioannidis J. P. A. *Empirical Assessment of Published Ef-*

fect Sizes and Power in the Recent Cognitive Neuroscience and Psychology Literature. PLOS Biology. 15, no. 3 (2017): e2000797.

47 Nelson L. D. et al. *Psychology's Renaissance*. Annual Review of Psychology. 69, no. 1 (2018): 511–34.

48 Это своего рода “проклятие победителя”, иногда обсуждаемое на аукционах, — когда человек, сделавший выигрышную ставку, переоценивает свой лот, каким бы тот ни был. В науке это еще называют “феноменом Протея”, в честь героя греческой мифологии, принимавшего любое обличье. Идея в том, что на первых порах, когда некий эффект только-только открыли, его величина в разных исследованиях зачастую сильно различается, частично из-за тех проблем, которые мы обсуждали в контексте статистической мощности и неспособности некоторых исследований обнаруживать небольшие эффекты. См. Ioannidis J. P. A., Trikalinos T. A. *Early Extreme Contradictory Estimates May Appear in Published Research: The Proteus Phenomenon in Molecular Genetics Research and Randomized Trials*. Journal of Clinical Epidemiology. 58, no. 6 (2005): 543–9; Lemoine N. P. et al. *Underappreciated Problems of Low Replication in Ecological Field Studies.* Ecology. 97, no. 10 (2016): 2554–61; Button K. S. et al. *Power Failure.*

49 Похожая проблема затрагивает статьи, которые я цитировал выше, где рассматривалась статистическая мощность в конкретных областях. Статистическая мощность исследований оценивалась там задним числом: задавался вопрос “Какой она у них была, раз позволила обнаружить найденный ими эффект?”. Но если в тех исследованиях величина истинного эффекта переоценивалась, такой апостериорный метод завышает оценку мощности. Таким образом, проверка статистической мощности задним числом может внушить уверенность, будто мощность в вашем исследовании совершенно нормальна, хотя это не так. Лучше взять идеальную величину эффекта — ту, которую, исходя из практического значения вашего эффекта, вы бы сочли малой, средней или большой (например, используя значимую разницу по более добротному показателю, такому как изменение по шкале боли, доход в долларах, температура или скорость), — и сделать мощность вашего исследования достаточной (то есть включить в него достаточное количество участников или наблюдений), чтобы надежно этот эффект обнаружить. См. Gelman A. *Don't Calculate Post-Hoc Power Using Observed Estimate of Effect Size.* 2018. *www.stat.columbia.edu/~gelman/research/unpublished/power_surgery.pdf*

50 Lamberink H. J. et al. *Statistical Power of Clinical Trials*. Величина эффекта, о которой я здесь говорю, — это *d*-значение Коэна, равное 0,21. Интерпретацию с количеством людей, которым станет лучше от лечения, я дал с помощью очень полезного калькуля-

тора с сайта, созданного Кристоффером Магнуссоном: *https://rpsychologist.com/d3/cohend.*

51 Leucht S. et al. *How Effective Are Common Medications: A Perspective Based on Meta-Analyses of Major Drugs*. BMC Medicine. 13, no. 1 (2015): 253. В этом исследовании, посвященном величине эффекта от часто используемых медицинских препаратов, отмечены некоторые широко распространенные лекарства, оказывающие большой эффект (например, ингибиторы протонного насоса вроде омепразола, очень сильно влияющие на секрецию желудочного сока), но и несколько таких, чьи эффекты удивительно малы (скажем, аспирин для профилактики сердечно-сосудистых заболеваний). Разумеется, даже лекарства, оказывающие скромный эффект, могут приносить огромную пользу в масштабах общества — позволяя экономить громадные деньги на медицинском обслуживании, когда такие лекарства прописывают миллионам людей, кому они нужны. Тем не менее авторы исследования считают, что “нам стоит адекватнее оценивать эффективность лекарств”. Про величину эффекта для трех препаратов, упомянутых мною в тексте, сказано, что ей соответствует *d*-значение Коэна, равное 0,55. См. также Pereira T. V. et al. *Empirical Evaluation of Very Large Treatment Effects of Medical Interventions*. JAMA. 308, no. 16 (2012): 1676–84.

52 См., например: Gignac G. E., Szodorai E. T. *Effect Size Guidelines for Individual Differences Researchers*. Personality and Individual Differences. 102 (2016): 74–8.

53 Благо, со времен Сирила Бёрта проведено было множество надежных близнецовых исследований. См. обзор: Polderman T. J. C. et al. *Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies*. Nature Genetics. 47, no. 7 (2015): 702–9.

54 О связи генов-кандидатов с когнитивными способностями см. обзор: Payton A. *The Impact of Genetic Research on Our Understanding of Normal Cognitive Ageing: 1995 to 2009*. Neuropsychology Review. 19, no. 4 (2009): 451–77.

55 De Quervain D. J.-F. et al. *A Functional Genetic Variation of the 5-HT2a Receptor Affects Human Memory*. Nature Neuroscience. 6, no. 11 (2003): 1141–2.

56 Munafo M. R. et al. *Serotonin Transporter (5-HTTLPR) Genotype and Amygdala Activation: A Meta-Analysis*. Biological Psychiatry. 63, no. 9 (2008): 852–7.

57 Сегодня фунтов за сто вы можете отослать пробу слюны напрямую компании, занимающейся генотипированием, и через несколько недель узнать, какие генетические варианты есть в вашем геноме.

58 Возможно, вы задаетесь вопросом, не страдает ли полногеномный поиск ассоциаций, раз проверяется потенциальная связь какого-то признака со многими тысячами генетических вариантов,

от проблемы множественных сравнений, которую мы обсуждали в предыдущей главе, — когда вместе с количеством вычисляемых *p*-значений возрастает риск ложноположительного результата. Исследователи, пользующиеся этим методом, хорошо осведомлены об опасности — и соответствующим образом резко понизили критерий для *p*-значений. Отказавшись от порога в 0,05, они считают значимыми только те *p*-значения, которые оказываются меньше 5×10^{-8} (или 0,00000005).

59 Duncan L. E. et al. *How Genome-Wide Association Studies (GWAS) Made Traditional Candidate Gene Studies Obsolete*. Neuropsychopharmacology. 44, no. 9 (2019): 1518–23.

60 Слово “редкие” здесь ключевое. Так, ученым известно о многих редких мутациях, связанных с пониженной обучаемостью и некоторыми формами расстройств аутистического спектра. Например, см. Niemi M. E. K. et al. *Common Genetic Variants Contribute to Risk of Rare Severe Neurodevelopmental Disorders*. Nature. 562, no. 7726 (2018): 268–71. Насколько мне известно, единственный наиболее распространенный “ген-кандидат”, выстоявший под натиском полногеномного поиска ассоциаций, — это вариация в гене *APOE*, которая, похоже, действительно надежно связана с риском развития болезни Альцгеймера. См. Marioni R. E. et al. *GWAS on Family History of Alzheimer’s Disease*. Translational Psychiatry. 8, no. 1 (2018): 99.

61 Коэффициент интеллекта: Chabris C. F. et al. *Most Reported Genetic Associations with General Intelligence Are Probably False Positives*. Psychological Science. 23, no. 11 (2012): 1314–23. Депрессия: Border R. et al. *No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples.* American Journal of Psychiatry. 176, no. 5 (2019): 376–87. Шизофрения: Farrell M. S. et al. *Evaluating Historical Candidate Genes for Schizophrenia.* Molecular Psychiatry. 20, no. 5 (2015): 555–62.

62 Alexander S. *5-HTTLPR: A Pointed Review.* Slate Star Codex. 7 May 2019.

63 Например, низкая статистическая мощность упоминается в аннотации этой статьи: Clarke H. et al. *Association of the 5-HTTLPR Genotype and Unipolar Depression: A Meta-Analysis*. Psychological Medicine. 40, no. 11 (2010): 1767–78. Кстати, готов поспорить, что в статьях о генах-кандидатах просто царило публикационное смещение. О некоторых свидетельствах в пользу моей гипотезы применительно к исследованиям взаимодействия генов-кандидатов с окружающей средой см. Duncan L. E., Keller M. C. *A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry*. American Journal of Psychiatry. 168, no. 10 (2011): 1041–9.

64 Fisher R. A. *XV. — The Correlation between Relatives on the Supposition of Mendelian Inheritance*. Transactions of the Royal Society of Edinburgh. 52, no. 2 (1919): 399–433. См. историческую дискуссию: Visscher P. M. et al. *10 Years of GWAS Discovery: Biology, Function, and Translation*. American Journal of Human Genetics. 101, no. 1 (2017): 5–22.

65 Что не означает, будто генетики близки к полному пониманию генетических основ сложных признаков. Они просто отбросили ущербный старый метод и заменили его на методы получше. Впереди еще громадный объем работы по выяснению того, какие гены вовлечены, как находить их среди различных групп людей, как наш анализ может расстроиться из-за трудностей, связанных с социальными или демографическими различиями, как именно гены оказывают свое влияние, что нам делать с нашими знаниями о генетике в контексте помощи людям, страдающим от заболеваний, и так далее. См. полезный обзор: Tam V. et al. *Benefits and Limitations of Genome-Wide Association Studies*. Nature Reviews Genetics. 20, no. 8 (2019): 467–84.

66 В исследованиях на животных существует интересный парадокс статистической мощности. Это может показаться нелогичным, но нам, вероятно, лучше использовать *побольше* животных в своих исследованиях, по крайней мере в краткосрочной перспективе. Увеличение размера выборки, а следовательно, и статистической мощности означает, что мы будем получать более надежные результаты — которые выдержат проверку временем и воспроизведением — и избежим длительных циклов неоднозначных, не дающих ничего нового экспериментов, а вероятно, и гибели большего числа животных в долгосрочной перспективе.

67 Например, см. Critchley C. R. *Public Opinion and Trust in Scientists: The Role of the Research Context, and the Perceived Motivation of Stem Cell Researchers.* Public Understanding of Science. 17, no. 3 (2008): 309–27.

Глава 6. **Хайп**

1 Harrison R. *The Springfield Files // The Simpsons.* Moore S. D. (dir.) Season 8. Episode 10. 12 Jan. 1997.

2 Wolfe-Simon F. et al. *A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus*. Science. 332, no. 6034 (2011): 1163–6.

3 Строго говоря, это не сталагмиты, а *туф* — он выглядит похоже, но обладает немного другими свойствами, например, он пористый внутри. *https://itotd.com/articles/2773/tufa/*

4 Исследование Волф-Саймон относится к астробиологии. Настоящей инопланетной жизни для изучения в лаборатории у астро-

биологов нет, поэтому они изучают, какой та *могла бы* быть. Один из способов — исследовать так называемых экстремофилов, например бактерии, живущие в суровых условиях, вроде таких, как в озере Моно.

5 Davies P. *The "Give Me a Job" Microbe*. Wall Street Journal. 4 Dec. 2010.

6 Цит. по: Clynes T. *Scientist in a Strange Land.* Popular Science. 26 Sept. 2011. Еще у Волф-Саймон брали интервью для журнала *Glamour*: Gowen A. *This Rising Star's Four Rules for You*. Glamour. June 2011.

7 В статье научного журналиста Карла Циммера приведено много цитат скептиков: Zimmer C. *"This Paper Should Not Have Been Published": Scientists See Fatal Flaws in the NASA Study of Arsenic-Based Life*. Slate. 7 Dec. 2010.

8 Многие сообщения в блоге Редфилд можно найти по хештегу *#arseniclife*: *http://rrresearch.fieldofscience.com.*

9 Editorial. *Response Required.* Nature. 468, no. 7326 (2010): 867.

10 Пресс-релиз можно найти по этой ссылке: *www.nasa.gov/home/hqnews/2010/nov/HQ_M10-167_Astrobiology.html.*

11 Kottke J. *Has NASA Discovered Extraterrestrial Life?* Kottke. 29 Nov. 2010. См. также статью с кадром из фильма "Инопланетянин" *(E. T. The Extra-Terrestrial): NASA to Unveil Details of Quest for Alien Life*. Fox News. 2 Dec. 2010.

12 Цит. по: Phillips T. (ed.) *Discovery of "Arsenic-Bug" Expands Definition of Life*. 2 Dec. 2010.

13 Все они перечислены в заметке Брюса Альбертса, тогдашнего главного редактора журнала *Science*: Alberts B. *Editor's Note.* Science. 332, no. 6034 (2011): 1149.

14 Reaves M. L. et al. *Absence of Detectable Arsenate in DNA from Arsenate-Grown GFAJ-1 Cells.* Science. 337, no. 6093 (2012): 470–3.

15 Erb T. J. et al. *GFAJ-1 Is an Arsenate-Resistant, Phosphate-Dependent Organism*. Science. 337, no. 6093 (2012): 467–70.

16 Clynes T. *Scientist in a Strange Land.*

17 Но это не значит, что такого никогда не происходит. Профессиональный риск, связанный с публикацией научных работ, заключается в том, что ваши результаты в средствах массовой информации могут быть искажены, неверно истолкованы или исковерканы. Ошибки бывают всякими — от несущественных до серьезных и вредных. К последним относится следующий случай: паникерство в медиа в 2011 году по поводу побочных эффектов статинов — хорошо зарекомендовавших себя и безопасных препаратов, снижающих риск сердечно-сосудистых заболеваний, — похоже, привело к тому, что люди в последующие годы перестали их принимать. Matthews A. et al. *Impact of Statin Related Media Coverage on Use of Statins: Interrupted Time Series Analysis with UK Primary Care Data*. BMJ. 353 (2016): i3283.

18 Sumner P. S. et al. *The Association between Exaggeration in Health-Related Science News and Academic Press Releases: Retrospective Observational Study*. BMJ. 349 (2014): g7015.

19 Чудесное краткое изложение истории моделирования на животных: Ericsson A. C. et al. *A Brief History of Animal Modeling*. Missouri Medicine. 110, no. 3 (2013): 201–5.

20 Contopoulos-Ioannidis D. G. et al. *Life Cycle of Translational Research for Medical Interventions*. Science. 321, no. 5894 (2008): 1298–9.

21 Garner J. P. *The Significance of Meaning: Why Do Over 90% of Behavioral Neuroscience Results Fail to Translate to Humans and What Can We Do to Fix It?* ILAR Journal. 55, no. 3 (2014): 438–56.

22 *https://twitter.com/justsaysinmice.* Кстати, не стоит забывать о низкокачественности значительного числа исследований на животных вроде мышей, что мы обсуждали в предыдущей главе.

23 Часто встречается немного иная формулировка: “корреляция не *подразумевает* причинно-следственной связи”. Здесь из-за многозначности слова “подразумевать” возникает некая двусмысленность. В строгом смысле этого слова (из А логически следует Б, как, например, существование танца *подразумевает*, что есть и танцор) — все правильно. Но в нестрогом смысле (А предполагает Б, но явным образом не называет, как получение по электронной почте резковатого письма от вашего начальника может *подразумевать*, что он вами недоволен) — формулировка неверна. В этом нестрогом смысле корреляция иногда действительно подразумевает причинно-следственную связь, даже если той вовсе и нет. Скажем так: если бы корреляция не подразумевала причинно-следственную связь в нестрогом смысле слова “подразумевать”, между двумя этими случаями не было бы такой путаницы.

24 Corley J. et al. *Caffeine Consumption and Cognitive Function at Age 70: The Lothian Birth Cohort 1936 Study*. Psychosomatic Medicine. 72, no. 2 (2010): 206–14.

25 Есть еще одна, менее известная причина, по которой две переменные могут коррелировать, — “ошибка коллайдера”. Вот здесь автор одного блога прекрасно все объясняет: Rohrer J. *That One Weird Third Variable Problem Nobody Ever Mentions: Conditioning on a Collider*. The 100% CI. 14 March 2017. В приведенном там примере, если мы просто посмотрим на выборку студентов колледжа, мы, к своему удивлению, можем увидеть отрицательную корреляцию между коэффициентом интеллекта и добросовестностью, хотя в популяции в целом такой корреляции вовсе и нет. А все потому, что и коэффициент интеллекта, и добросовестность повышают шансы в принципе поступить в колледж, отбраковывая тех людей, у кого оба этих показателя низки. То обстоятельство, что в выборке студентов колледжа отсутствуют люди с низким интеллектом

и низкой добросовестностью, порождает иллюзорную корреляцию между двумя этими переменными. Это коварная проблема, и распространена она в исследованиях шире, чем нам хотелось бы думать. Также см. Munafo M. R. et al. *Collider Scope: When Selection Bias Can Substantially Influence Observed Associations*. International Journal of Epidemiology. 47, no. 1 (2017): 226–35.

26 Если вы хотите окончательно себя растревожить, почитайте, что философ Дэвид Юм писал о так называемой проблеме индукции, которая, по сути, гласит, что корреляция — это даже не корреляция, поскольку нет рациональных оснований утверждать, что события, произошедшие в прошлом, произойдут снова. (Классическая задачка: докажите, что солнце взойдет завтра утром; "Так всегда было раньше" — на самом деле не логическое основание для веры в завтрашний восход.) Эта проблема обсуждается в философии уже сотни лет, и, хотя немало чрезвычайно умных мыслителей пытались ее разрешить, многие думают, что она абсолютно неразрешима. Замечательное обсуждение см. тут: Henderson L. *The Problem of Induction*. Stanford Encyclopedia of Philosophy. Winter 2019.

27 Adams R. C. et al. *Claims of Causality in Health News: A Randomised Trial*. BMC Medicine. 17, no. 1 (2019): 91.

28 Boutron I. et al. *Three Randomized Controlled Trials Evaluating the Impact of "Spin" in Health News Stories Reporting Studies of Pharmacologic Treatments on Patients'/Caregivers' Interpretation of Treatment Benefit*. BMC Medicine. 17, no. 1 (2019): 105.

29 Davies N. *Flat Earth News: An Award-Winning Reporter Exposes Falsehood, Distortion and Propaganda in the Global Media*. London: Vintage Books, 2009; Jackson D., Moloney K. *Inside Churnalism: PR, Journalism and Power Relationships in Flux*. Journalism Studies. 17, no. 6 (2016): 763–80.

30 Dumas-Mallet E. et al. *Poor Replication Validity of Biomedical Association Studies Reported by Newspapers*. PLOS ONE. 12, no. 2 (2017): e0172650.

31 Это не значит, что научно-популярные книги, написанные неспециалистами, нужно сбросить со счетов: они также подвержены серьезным проблемам. В своей рецензии на сборник эссе Малкольма Гладуэлла "Что видела собака" Стивен Пинкер ввел термин "проблема значений айгона" *(Igon Values Problem)* для описания случая, когда Гладуэлл исказил понятие "собственные значения" (*eigenvalues;* математическое понятие, важное для статистического анализа) — скорее всего, он услышал, как один из его собеседников произнес это, а затем не удосужился поискать нужное слово для проверки. [Отметим, что в опубликованном переводе этой книги на русский язык, выполненном Е. Бакушевой, словосочетание *igon value* вообще пропало. *(Прим. перев.)*] "Значения айгона"

частенько встречаются в научно-популярной литературе, обнажая пробелы в понимании, которые могут возникнуть, если автор не является экспертом в рассматриваемом вопросе. Однако, как мы вскоре увидим, и ученые, даже пишущие о тех областях, которыми *сами* занимаются, способны создавать книги с проблемами ничуть не лучше "значений айгона". Pinker S. *Malcolm Gladwell, Eclectic Detective*. New York Times. 7 Nov. 2009.

32 Dweck C. S. *Mindset: The New Psychology of Success.* New York: Ballantine Books, 2008. Еще у Двек есть очень популярное выступление на TEDx, с общим количеством просмотров 13,5 миллиона на момент написания этих строк (10,2 миллиона на сайте TED и 3,3 миллиона на *YouTube*), где она заявляет, что это "базовое человеческое право детей, всех детей, — жить там, где поощряется развитие". Dweck C. *The Power of Believing That You Can Improve*. TEDxNorrkoping. Nov. 2014.

33 Там же.

34 Yettick H. et al. *Mindset in the Classroom: A National Study of K-12 Teachers.* Editorial Projects in Education. Education Week Research Center, 2016. Обратите внимание: выборка нерепрезентативна, так что будьте бдительны. В феврале 2020 года *Google* при поиске по слову *mindset* ("образ мышления") только в доменах "*.sch.uk*", то есть исключительно на сайтах школ Великобритании, показывал 43 200 результатов. Согласитесь, это дает некоторое представление о популярности идеи. Сейчас в Великобритании чуть больше 32 тысяч школ. *Key UK Education Statistics.* British Educational Suppliers Association. 28 Oct. 2019.

35 Sisk V. F. et al. *To What Extent and Under Which Circumstances Are Growth Mind-Sets Important to Academic Achievement? Two Meta-Analyses*. Psychological Science. 29, no. 4 (2018): 549–71. О дальнейшей работе той же группы исследователей, скептически относящихся к установкам на развитие, см. Burgoyne A. P. et al. *How Firm Are the Foundations of Mind-Set Theory? The Claims Appear Stronger Than the Evidence*. Psychological Science. 31, no. 3 (2020): 258–267.

36 Для фанатов статистики: корреляционный эффект соответствовал коэффициенту корреляции Пирсона, равному 0,10, а экспериментальный эффект — *d*-значению Коэна, равному 0,08. Представить себе, что означает *d*-значение 0,08 помимо перекрывания распределений на 96,8 %, можно так: допустим, мы наугад выбираем человека из группы, прошедшей тренинг, и смотрим, выше или ниже его отметки по отношению к среднему в контрольной группе. Если бы эффекта от тренинга не было, вероятность такого отклонения равнялась бы 50 % (средние одинаковы). Метаанализ показал, что эффект от тренинга такой, как если бы студенты, прошедшие

обучение по формированию подвижного образа мышления, с вероятностью 52,3 % имели отметки выше среднего в контрольной группе. Вычислено здесь: *https://rpsychologist.com/d3/cohend.*

37 В метаанализе обнаружились некоторые свидетельства того, что для детей, находящихся в группе риска (из бедных семей, например), польза от тренингов по подвижному образу мышления больше. Это подтвердилось в недавнем масштабном исследовании, проведенном сторонниками установок на развитие, результаты которого оказались в целом схожи с результатами метаанализа. Yeager D. S. et al. *A National Experiment Reveals Where a Growth Mindset Improves Achievement*. Nature. 573, no. 7774 (2019): 364–9.

38 Если вам интересна наглядная иллюстрация того, как хайп провоцирует ощущение научной "ползучей миссии", прочитайте статью, опубликованную Двек и ее коллегами в журнале *Science* в 2011 году, где выдвинуто предположение — на основании довольно хило выглядящих доказательств, — что концепцию подвижного образа мышления можно использовать для содействия миру на Ближнем Востоке. Halperin E. et al. *Promoting the Middle East Peace Process by Changing Beliefs About Group Malleability.* Science. 333, no. 6050 (2011): 1767–9.

39 Другой пример хайпа в образовании до некоторой степени вытекает из концепции подвижного образа мышления: это "упорство". Психолог Анжела Дакворт продвигала идею, что способность настойчиво заниматься увлекшей вас задачей и не сдаваться, даже когда жизнь чинит препятствия, — это ключ к жизненному успеху, причем способность эта куда важнее, чем врожденный талант. Идея стала крайне популярной: на момент написания этих строк выступление Дакворт на конференции TED на эту тему набрало 25,5 миллиона просмотров (19,5 миллиона на сайте TED и еще 6 миллионов на *YouTube*; Duckworth A. L. *Grit: The Power of Passion and Perseverance*. TED Talks Education. April 2013), а вышедшая затем книга "Упорство: сила увлеченности и настойчивости" *(Grit: The Power of Passion and Perseverance)* стала бестселлером по версии *The New York Times* и продолжает стабильно продаваться. Как и концепция образа мышления, идея упорства стала частью философии многих школ, включая чартерные школы KIPP (*Knowledge Is Power Program*, программа "Знание — сила"), крупнейшую в США группу чартерных образовательных учреждений, где обучаются почти 90 тысяч учеников (*www.kipp.org/approach/character/).* К чести Дакворт нужно заметить, что она обеспокоилась тем, насколько ее результаты оказались раздуты. В 2015 году она сказала журналисту из NPR, что "энтузиазм опережает науку" (Kamenetz A. *A Key Researcher Says "Grit" Isn't Ready For High-Stakes Measures.* NPR.

13 May 2015). Мудрое изречение, учитывая, что влияние упорства (или тренингов по его развитию) оценивается в метаанализах как чрезвычайно слабое. См. Credé M. et al. *Much Ado about Grit: A Meta-Analytic Synthesis of the Grit Literature*. Journal of Personality and Social Psychology. 113, no. 3 (2017): 492–511; Credé M. *What Shall We Do About Grit? A Critical Review of What We Know and What We Don't Know.* Educational Researcher. 47, no. 9 (2018): 606–11.

40 И вообще, именно в ответ на неудавшуюся попытку воспроизвести один из результатов Барга (обсуждавшееся нами во второй главе исследование про прайминг мыслями о пожилых, якобы заставлявший людей идти медленнее) Даниэль Канеман написал свое открытое письмо к социальным психологам, сообщая, что узрел "угрозу катастрофы", и призывая их изменить подход к исследованиям.

41 Bargh J. *Before You Know It: The Unconscious Reasons We Do What We Do*. London: Windmill Books, 2018.

42 Психолог Ульрих Шиммак опубликовал в своем блоге подробную "количественную рецензию" на книгу Барга, проанализировав, насколько можно доверять каждой из цитируемых Баргом статей. Schimmack U. *"Before You Know It" by John A. Bargh: A Quantitative Book Review.* Replication Index. 28 Nov. 2017. Кстати, "количественная рецензия на книгу" — прекрасная идея, которую многим другим ученым стоит взять на вооружение.

43 Bargh J. *Before You Know It*.

44 Chen S. et al. *Relationship Orientation as a Moderator of the Effects of Social Power*. Journal of Personality and Social Psychology. 80, no. 2 (2001): 173–87.

45 Chabris C. F. et al. *No Evidence That Experiencing Physical Warmth Promotes Interpersonal Warmth: Two Failures to Replicate*. Social Psychology. 50, no. 2 (2019): 127–32. Справедливости ради отметим, что конкретно эта статья о попытке повторить исследование с чашками кофе была опубликована уже после выхода книги Барга. Но поскольку другие его очень похожие исследования понятия "теплоты" (там вместо чашки кофе участники держали медицинскую грелку) тоже не воспроизвелись, думается, ему стоило меньше полагаться на те результаты. См. Lynott D. et al. *Replication of "Experiencing Physical Warmth Promotes Interpersonal Warmth" by Williams and Bargh (2008).* Social Psychology. 45, no. 3 (2014): 216–22.

46 Еще можно вспомнить историю Эми Кадди, чья книга-бестселлер была основана на результатах исследования, которое, как позже выяснилось, является образцовым примером *p*-хакинга. См. вторую и четвертую главы.

47 Walker M. *Why We Sleep: The New Science of Sleep and Dreams*. London: Allen Lane, 2017. [Уолкер М. *Зачем мы спим. Новая наука о сне и сновидениях*. М.: КоЛибри, 2018.]

48 Walker M. *Sleep Is Your Superpower*. TED. April 2019. Указанное количество просмотров — это сумма 6,7 миллиона просмотров на сайте TED и еще 3,3 миллиона на *YouTube* (по состоянию на ноябрь 2019 года).

49 Smith R. *Why We Sleep — One of Those Rare Books That Changes Your Worldview and Should Change Society and Medicine*. The BMJ Opinion. 19 June 2018.

50 Walker M. *Why We Sleep.*

51 Guzey A. *Matthew Walker's "Why We Sleep" Is Riddled with Scientific and Factual Errors*. 15 Nov. 2019.

52 Shen X. et al. *Nighttime Sleep Duration, 24-Hour Sleep Duration and Risk of All-Cause Mortality among Adults: A Meta-Analysis of Prospective Cohort Studies*. Scientific Reports. 6, no. 1 (2016): 21480.

53 Chen Y. et al. *Sleep Duration and the Risk of Cancer: A Systematic Review and Meta-Analysis Including Dose-Response Relationship*. BMC Cancer. 18, no. 1 (2018): 1149.

54 Gelman A. *"Why We Sleep" Data Manipulation: A Smoking Gun? Statistical Modeling, Causal Inference, and Social Science*. 27 Dec. 2019. Обратите внимание, что существует и ответ на некоторые критические замечания к книге "Зачем мы спим", вроде бы от Уолкера: *Why We Sleep: Responses to Questions from Readers.* SleepDiplomat. 19 Dec. 2019.

55 Некоторые размышления о научно-популярном писательстве см. тут: Chabris C. F. *What Has Been Forgotten About Jonah Lehrer*. 12 Feb. 2013. *http://blog.chabris.com/2013/02/what-has-been-forgotten-about-jonah.html*

56 Vinkers C. H. et al. *Use of Positive and Negative Words in Scientific PubMed Abstracts between 1974 and 2014: Retrospective Analysis*. BMJ. 351 (2015): h6467. Обратите внимание: анализ скорректирован с учетом растущего числа ежегодно публикуемых статей. Определенные фразы заслуживают того, чтобы проанализировать и их тоже: "Я потерял счет случаям, когда я читал (а иногда и писал!) «в этой работе впервые рассматривается…»"

57 Позднее специально изучалось, насколько часто слово "беспрецедентный" используется в области раковых исследований. Выяснилось, что примерно в трети случаев оно применялось неправомерно: хотя авторы заявляли, что их результат "беспрецедентен", в какой-то более ранней статье о том же виде лечения сообщалось про еще больший эффект. См. Duggan K. T. et al. *Use of Word "Unprecedented" in the Media Coverage of Cancer Drugs: Do "Unprecedented" Drugs Live up to the Hype?* Journal of Cancer Policy. 14 (2017): 16–20.

58 Vinkers C. H. et al. *Use of Positive and Negative Words*. Как ни странно, частота использования негативно звучащих слов в ан-

нотациях тоже увеличилась, хотя и незначительно. Возможно, правильнее будет сказать, что ученые стали писать более "яркие" аннотации. Частота использования нейтральных и произвольно выбранных слов не изменилась вообще.

59 На самом деле есть некоторые свидетельства того, что научный прогресс со временем замедляется. Cowen T., Southwood B. *Is the Rate of Scientific Progress Slowing Down?* SSRN. 2019.

60 Nature: *www.nature.com/authors/author_resources/about_npg.html;* Science: *www.sciencemag.org/about/mission-and-scope;* Cell: *www.cell.com/cell/aims;* Proceedings of the National Academy of Sciences: *www.pnas.org/page/authors/purpose-scope.*

61 New England Journal of Medicine: *www.nejm.org/about-nejm/about-nejm.*

62 Boutron I. *Reporting and Interpretation of Randomized Controlled Trials with Statistically Nonsignificant Results for Primary Outcomes*. JAMA. 303, no. 20 (2010): 2058–64. Также см. Boutron I., Ravaud P. *Misrepresentation and Distortion of Research in Biomedical Literature*. Proceedings of the National Academy of Sciences. 115, no. 11 (2018): 2613–9.

63 Hankins M. *Still Not Significant*. Probable Error. 21 April 2013. О преобладании подобных утверждений в онкологической литературе см. Nead K. T. et al. *The Use of "Trend" Statements to Describe Statistically Nonsignificant Results in the Oncology Literature*. JAMA Oncology. 4, no. 12 (2018): 1778–9. Было отмечено, что подобные утверждения пишутся так, словно результат всегда движется *по направлению* к статистической значимости, но раз уж вы собираетесь думать о своих результатах в таком странном телеологическом ключе, не учитывая сути *p*-значений, то откуда вам знать, что цифры не уходят *от* значимости? Коли на то пошло, откуда вы знаете, что те ваши результаты, которые находятся на предпочтительной стороне от порога статистической значимости, не стремятся забраться выше 0,05? По загадочным причинам ученые, похоже, никогда не чувствуют необходимости описывать *p*-значения, которые чуть *меньше* 0,05, как "тенденцию в направлении *от* статистической значимости" (см. раздел *Dredging for P* на сайте *www.senns.uk/wprose.html).* Далее, может показаться, что я сам себе противоречу, описав в четвертой главе порог для *p*-значений как произвольный, но при этом критикуя тех ученых, которые получают результат немного выше порога и все равно интерпретируют его как значимый (или в каком-то смысле приближающийся к значимости). Суть в том, что если вы вступаете в игру с *p*-значениями, то должны придерживаться правил. Если вы исходно заявили, что будете считать статистически значимыми только результаты с $p < 0{,}05$, то не можете после получения результатов просто по-

менять правила игры. Иначе порог теряет свою единственную полезную функцию — контролировать риск ложноположительного результата при ошибочности проверяемой гипотезы.

64 Turrentine M. *It's All How You "Spin" It: Interpretive Bias in Research Findings in the Obstetrics and Gynecology Literature*. Obstetrics & Gynecology. 129, no. 2 (2017): 239–42.

65 Kempf E. et al. *Overinterpretation and Misreporting of Prognostic Factor Studies in Oncology: A Systematic Review*. British Journal of Cancer. 119, no. 10 (2018): 1288–96. В тридцати одной работе из этого обзора подача под нужным углом выглядела, например, так: ученые прятали *p*-значения для своих результатов, не являющихся статистически значимыми, в сноску под таблицей с числами, однако же демонстративно объявляли о *p*-значениях для результатов значимых в самой таблице.

66 Austin J. et al. *Evaluation of Spin within Abstracts in Obesity Randomized Clinical Trials: A Cross-Sectional Review*. Clinical Obesity. 9, no. 2 (2019): e12292.

67 Beijers L. et al. *Spin in RCTs of Anxiety Medication with a Positive Primary Outcome: A Comparison of Concerns Expressed by the US FDA and in the Published Literature*. BMJ Open. 7, no. 3 (2017): e012886.

68 Mehler D. M. A., Kording K. P. *The lure of misleading causal statements in functional connectivity research*. ArXiv (8 Dec. 2018): 1812.03363. В этой статье отмечается, что все чаще используется понятие "причинность по Грэнджеру". В 1960-х годах его ввел экономист Клайв Грэнджер, нобелевский лауреат. Идея в том, что если данные из одного "временного ряда" (скажем, колебания на фондовом рынке) предсказывают последующие изменения в другом (например, изменения других экономических показателей страны), то это уже немного выходит за пределы обычной корреляции. В таком случае некоторые исследователи говорят, что колебания на фондовом рынке "вызвали по Грэнджеру" изменения в экономике. Подобные корреляции интересны, но это все равно лишь корреляции, поэтому обычные опасения насчет третьей, спутывающей переменной применимы здесь в той же степени (то есть третий, параллельный временной тренд мог вызвать как исходные колебания рынка, так и последующие изменения экономических показателей). Использовать слово "вызвали", когда ваше исследование не предназначено для выявления причинно-следственных связей (с помощью эксперимента или другого хитроумного способа обнаружить причинно-следственную структуру данных), — значит играть с огнем.

69 Wu T. et al. *Randomized Trials Published in Some Chinese Journals: How Many Are Randomized?* Trials. 10, no. 1 (2009): 46.

70 McGrath T. A. et al. *Overinterpretation of Research Findings: Evidence of "Spin" in Systematic Reviews of Diagnostic Accuracy Studies*.

Clinical Chemistry. 63, no. 8 (2017): 1362. Также см. Chiu K. et al. *"Spin" in Published Biomedical Literature: A Methodological Systematic Review*. PLOS Biology. 15, no. 9 (2017): e2002173.

71 Насколько мне известно, было проведено одно рандомизированное контролируемое исследование, когда в клиническом контексте изучалось влияние подачи результатов в научных статьях под нужным углом. Ученые взяли подборку аннотаций из статей, посвященных онкологическим исследованиям, где сообщалось об отрицательных результатах, но поданных так, чтобы они казались положительными, и переписали их без всяких хитростей, убрав все неоправданные преувеличения и убедившись, что результаты представлены честно. И затем показали эти аннотации тремстам клиницистам — в конце концов, они и есть целевая аудитория, поскольку регулярно принимают решения по выбору лекарств и других методов лечения. Как и следовало ожидать, клиницисты сочли те методы лечения, о которых говорилось в "раскрученных" аннотациях, более действенными. Важно, правда, отметить, что эффекты в этой работе были довольно небольшими, с *p*-значениями чуть ниже порога в 0,05. Хотелось бы увидеть повторение исследования, прежде чем слишком доверять его результатам, — вот почему я рассказываю о нем только тут, в примечании. Boutron I. et al. *Impact of Spin in the Abstracts of Articles Reporting Results of Randomized Controlled Trials in the Field of Cancer: The SPIIN Randomized Controlled Trial.* Journal of Clinical Oncology. 32, no. 36 (2014): 4120–6.

72 Yong E. *I Contain Multitudes: The Microbes within Us and a Grander View of Life*. New York: HarperCollins, 2016. [Йонг Э. *Как микробы управляют нами. Тайные властители жизни на Земле*. М.: АСТ, 2017.]

73 Caulfield T. *Microbiome Research Needs a Gut Check*. Globe and Mail. 11 Oct. 2019.

74 Shane A. L. *The Problem of DIY Fecal Transplants*. Atlantic. 16 July 2013.

75 Kao D. et al. *Effect of Oral Capsule- vs Colonoscopy-Delivered Fecal Microbiota Transplantation on Recurrent Clostridium Difficile Infection: A Randomized Clinical Trial*. JAMA. 318, no. 20 (2017): 1985. Впервые применение подобного трансплантата было описано в 1958 году, однако интереса к такому способу лечения много десятков лет не наблюдалось. Eiseman B. et al. *Fecal Enema as an Adjunct in the Treatment of Pseudomembranous Enterocolitis*. Surgery. 44, no. 5 (1958): 854–9.

76 Hui W. et al. *Fecal Microbiota Transplantation for Treatment of Recurrent C. Difficile Infection: An Updated Randomized Controlled Trial Meta-Analysis*. PLOS ONE. 14, no. 1 (2019): e0210016; Rokkas T. et al. *A Network Meta-Analysis of Randomized Controlled Trials Exploring the Role of Fecal Microbiota Transplantation in Recurrent Clostri-*

dium Difficile Infection. United European Gastroenterology Journal. 7, no. 8 (2019): 1051–63.

77 Микробиом и депрессия, тревожность и шизофрения: Foster J. A., Neufeld K.-A. M. *Gut — Brain Axis: How the Microbiome Influences Anxiety and Depression*. Trends in Neurosciences. 36, no. 5 (2013): 305–12; Dinan T. G. et al. *Genomics of Schizophrenia: Time to Consider the Gut Microbiome?* Molecular Psychiatry. 19, no. 12 (2014): 1252–7. Сердечно-сосудистые заболевания: Ahmadmehrabi S., Tang W. H. W. *Gut Microbiome and Its Role in Cardiovascular Diseases*. Current Opinion in Cardiology. 32, no. 6 (2017): 761–6. Ожирение: Marotz C. A., Zarrinpar A. *Treating Obesity and Metabolic Syndrome with Fecal Microbiota Transplantation.* Yale Journal of Biology and Medicine. 89, no. 3 (2016): 383–8. Рак: Chen D. et al. *Fecal Microbiota Transplantation in Cancer Management: Current Status and Perspectives*. International Journal of Cancer. 145, no. 8 (2019): 2021–31. Болезнь Альцгеймера: Sandoiu A. *Stool Transplants from "Super Donors" Could Be a Cure-All*. Medical News Today. 22 January 2019. Болезнь Паркинсона: Van Laar T. et al. *Faecal Transplantation, Pro- and Prebiotics in Parkinson's Disease; Hope or Hype?* Journal of Parkinson's Disease. 9, no. s2 (2019): S371–9. Аутизм: Bibbo S. et al. *Fecal Microbiota Transplantation: Past, Present and Future Perspectives.* Minerva Gastroenterologica e Dietologica. 63, no. 4 (2017): 420–30.

78 Я использовал в этом предложении слово "вызывая", однако причинно-следственный статус многих утверждений о микробиоме далеко не однозначен. См. Lynch K. E. et al. *How Causal Are Microbiomes? A Comparison with the Helicobacter Pylori Explanation of Ulcers*. Biology & Philosophy. 34, no. 6 (2019): 62.

79 Напомню, что вопрос о связи между здоровьем кишечника и аутизмом уже поднимался Эндрю Уэйкфилдом, о чем говорилось в третьей главе. Мы пока совершенно не понимаем, различия ли в микробиоме вызывают аутизм, или, наоборот, эти различия им обусловлены — например, из-за того что у людей с аутизмом часто очень ограниченный рацион.

80 Хотя где-то они использовали образцы, взятые только у пяти детей с аутизмом и трех "контрольных".

81 Sharon G. et al. *Human Gut Microbiota from Autism Spectrum Disorder Promote Behavioral Symptoms in Mice*. Cell. 177, no. 6 (2019): 1600–18.e17.

82 Lowe D. *Autism Mouse Models for the Microbiome?* In the Pipeline. 31 May 2019.

83 Sharon G. et al. *Human Gut Microbiota*.

84 *Gut Bacteria Influence Autism-like Behaviors in Mice*. California Institute of Technology. News release. 30 May 2019.

85 Brock J. *Can Gut Bacteria Cause Autism (in Mice)?* Medium. 14 June 2019. Также см. Zeliadt N. *Study of Microbiome's Importance in Autism Triggers Swift Backlash*. Spectrum News. 27 June 2019.

86 Lumley T. *Analysing the Mouse Microbiome Autism Data*. Not Stats Chat. 16 June 2019. Также см. собственный анализ Джона Брока: *https://rpubs.com/drbrocktagon/506022.*

87 Zheng P. et al. *The Gut Microbiome from Patients with Schizophrenia Modulates the Glutamate-Glutamine-GABA Cycle and Schizophrenia-Relevant Behaviors in Mice*. Science Advances. 5, no. 2 (2019): 8. Критику в форме дискуссии в твиттере см. тут: *https://twitter.com/WiringTheBrain/status/1095012297200844800.*

88 Есть еще и другие причины для скепсиса. В исследовании 2015 года, в котором участвовало свыше полутора тысяч человек, прошедших через колэктомию (хирургическое удаление толстой кишки, а значит, и всей ее микробиоты), через несколько лет после операции изучалось, реже ли у этих пациентов по сравнению с контрольными развивались сердечно-сосудистые заболевания. Если микробиом вносит весомый вклад в развитие этих болезней — а учитывая, что пациенты прошли через колэктомию, среди них были люди с очень нездоровым микробиомом, — мы вправе ожидать снижения риска сердечно-сосудистых заболеваний. Различий выявлено не было. Jensen A. B. et al. *Long-Term Risk of Cardiovascular and Cerebrovascular Disease after Removal of the Colonic Microbiota by Colectomy: A Cohort Study Based on the Danish National Patient Register from 1996 to 2014.* BMJ Open. 5, no. 12 (2015): e008702.

89 Hanage W. P. *Microbiology: Microbiome Science Needs a Healthy Dose of Scepticism*. Nature. 512, no. 7514 (2014): 247–8; Falony G. et al. *The Human Microbiome in Health and Disease: Hype or Hope*. Acta Clinica Belgica. 74, no. 2 (2019): 53–64; Taylor J. *The Microbiome and Mental Health: Hope or Hype?* Journal of Psychiatry and Neuroscience. 44, no. 4 (2019): 219–22.

90 “Улучшить ваши показатели”: Holtz A. *Harvard Researchers' Speculative, Poop-Based Sports Drink Company Raises Questions about Conflicts of Interest*. Health News Review. 19 Oct. 2017; также см. Editorial. *Probiotics: Elixir or Empty Promise?* Lancet Gastroenterology & Hepatology. 4, no. 2 (2019): 81. “Перфорация прямой кишки”: Shapiro N. *There Are Trillions Of Reasons Not To Cleanse Your Colon*. Forbes. 19 Sept. 2019; еще об опасностях см. Handley D. V. et al. *Rectal Perforation from Colonic Irrigation Administered by Alternative Practitioners*. Medical Journal of Australia. 181, no. 10 (2004): 575–6. В любом случае вся эта затея бессмысленна: через две недели после того, как микробы в ходе ирригации толстой кишки начисто вымываются, все возвращается на круги своя. См. Nagata N. et al. *Effects of Bowel Preparation on the Human Gut Microbiome and Meta-*

bolome. Scientific Reports. 9, no. 1 (2019): 4042. "Национальность вашего микробиома": *https://atlasbiomed.com/uk/microbiome/results.* Также см. Senapathy K. *Keep Calm And Avoid Microbiome Mayhem*. Forbes. 7 March 2016.

91 Молоко: Harkinson J. *The Scary New Science That Shows Milk Is Bad For You*. Mother Jones. Dec. 2015. Бекон: *Killer Full English: Bacon Ups Cancer Risk*. LBC News. 17 April 2019. Яйца: Physicians' Committee for Responsible Medicine. *New Study Finds Eggs Will Break Your Heart*. 16 March 2016. Подзаголовок был следующим: "Американцы едят по 279 яиц на человека ежегодно, и в новом исследовании выясняется, что это их **убивает** [жирный шрифт в оригинале]". Исходная работа: Zhong V. et al. *Associations of Dietary Cholesterol or Egg Consumption with Incident Cardiovascular Disease and Mortality*. JAMA. 321, no. 11 (2019): 1081. Подробный критический разбор: Rafi Z. *Revisiting Eggs and Dietary Cholesterol*. Less Likely. 22 March 2019. Кстати, вот тут высмеивается этот порок средств массовой информации (моя любимая пародия на эту тему): *Nutritional Shake-Up: The FDA Now Recommends That Americans Eat A Bowl Of 200 Eggs On Their 30th Birthday And Then Never Eat Any Eggs Again.* Clickhole. 24 Oct. 2017.

92 Эту мысль уже озвучивал исследователь-диетолог Кевин Клэтт: *https://twitter.com/kcklatt/status/902558341414694912.* Опрос, проведенный Британским фондом питания, показывает, что люди считают "противоречивые сообщения" об исследованиях, посвященных питанию, крайне неясными, хотя и непонятно, насколько выборка в том опросе была репрезентативной: *www.nutrition.org.uk/press-office/pressreleases/1156-mixedmessages.html.*

93 Подробно об истории Даса, умершего на следующий год после увольнения, рассказано здесь: Webb G. P. *Dipak Kumar Das (1946–2013) Who Faked Data about Resveratrol — the Magic Red Wine Ingredient That Cures Everything?* Dr Geoff Nutrition. 10 Nov. 2017.

94 Недавно разгорелся спор о качестве исследований, посвященных красному мясу: когда одна сторона была связана с мясной промышленностью, а другая — с компаниями, продающими вегетарианские продукты и оказывающимися в выигрыше, если люди сокращают потребление красного мяса. Rubin R. *Backlash Over Meat Dietary Recommendations Raises Questions About Corporate Ties to Nutrition Scientists*. JAMA. 323, no. 5 (2020): 401.

95 Ioannidis J. P. A., Trepanowski J. F. *Disclosures in Nutrition Research: Why It Is Different*. JAMA. 319, no. 6 (2018): 547.

96 См., например, сайт Национальной службы здравоохранения Англии (*www.nhs.uk/live-well/eat-well/different-fats-nutrition)* и вот эту статью от сотрудников американской клиники Мейо: Mayo Clinic Staff. *Dietary Fats: Know Which Types to Choose*. 1 Feb. 2019.

97 Hamley S. *The Effect of Replacing Saturated Fat with Mostly N-6 Polyunsaturated Fat on Coronary Heart Disease: A Meta-Analysis of Randomised Controlled Trials*. Nutrition Journal. 16, no. 1 (2017): 30.

98 Еще некоторым статьям, демонстрирующим отсутствие различий между типами жирных кислот, понадобилось аномально долгое время для публикации, что указывает на нежелание авторов — или рецензентов — видеть их напечатанными.

99 Полное объяснение см. здесь: Miettinen M. et al. *Effect of Cholesterol-lowering Diet on Mortality from Coronary Heart-Disease and Other Causes*. Lancet. 300, no. 7782 (1972): 835–8.

100 В любом эксперименте меняться должна только изучаемая переменная, здесь это тип потребляемых жирных кислот: насыщенные или ненасыщенные. Однако в некоторых исследованиях группы отличались еще и по другим параметрам: им, например, давали различные диетические рекомендации, а в одном случае в ходе больничного исследования участники принимали разные лекарства.

101 Schoenfeld J. D., Ioannidis J. P. A. *Is Everything We Eat Associated with Cancer? A Systematic Cookbook Review*. American Journal of Clinical Nutrition. 97, no. 1 (2013): 127–34.

102 И действительно, как и в других областях, в метаанализах, посвященных исходным исследованиям, в которых обнаружилось большое влияние на риск развития рака, эффекты обычно оказывались куда скромнее.

103 Список относящихся к питанию корреляций, которые в рандомизированных испытаниях не подтвердились, см. здесь: Young S. S., Karr A. *Deming, Data and Observational Studies: A Process out of Control and Needing Fixing*. Significance. 8, no. 3 (2011): 116–20. Впрочем, некоторое несогласие с этой идеей выражено тут (см. четвертый “миф”): Satija A. et al. *Perspective: Are Large, Simple Trials the Solution for Nutrition Research?* Advances in Nutrition. 9, no. 4 (2018): 381. Серьезная проблема заключается в том, что рандомизированные контролируемые испытания в силу своей дороговизны часто куда менее масштабны, чем наблюдательные исследования, поэтому сравнения некорректны, ведь это как сравнивать яблоки с апельсинами (хотя я уверен, что в анналах эпидемиологии питания имеется и такое исследование).

104 Westfall J., Yarkoni T. *Statistically Controlling for Confounding Constructs Is Harder than You Think*. PLOS ONE. 11, no. 3 (2016): e0152719.

105 Archer E. et al. *Controversy and Debate: Memory-Based Methods Paper 1: The Fatal Flaws of Food Frequency Questionnaires and Other Memory-Based Dietary Assessment Methods.* Journal of Clinical Epidemiology. 104 (2018): 113.

106 Суровую критику состоятельности опросников по частоте потребления различных пищевых продуктов см. здесь: Archer E. et al. *Validity of U. S. Nutritional Surveillance: National Health and Nutrition Examination Survey Caloric Energy Intake Data, 1971–2010.* PLOS ONE. 8, no. 10 (2013): e76632. Кто-то считает, что критика сама раздута и надуманна. Квалифицированную защиту см. тут: Hébert J. R. et al. *Considering the Value of Dietary Assessment Data in Informing Nutrition-Related Health Policy*. Advances in Nutrition. 5, no. 4 (2014): 447–55. Побольше об этих спорах можно почитать здесь: Berezow A. *Is Nutrition Science Mostly Junk?* American Council on Science and Health. 20 Nov. 2018; Nosowitz D. *The Bizarre Quest to Discredit America's Most Important Nutrition Survey*. TakePart. 29 July 2015.
107 Ioannidis J. P. A., Trepanowski J. F. *Disclosures in Nutrition Research.*
108 Satija A. et al. *Perspective*. Еще в этой статье авторы защищают исследования, посвященные питанию, от многих нападок, которые я привел выше в тексте. Также см. Giovannucci E. *Nutritional Epidemiology: Forest, Trees and Leaves*. European Journal of Epidemiology. 34, no. 4 (2019): 319–25. Ответ Иоаннидиса: Ioannidis J. P. A. *Unreformed Nutritional Epidemiology: A Lamp Post in the Dark Forest.* European Journal of Epidemiology. 34, no. 4 (2019): 327–31.
109 Estruch R. et al. *Primary Prevention of Cardiovascular Disease with a Mediterranean Diet*. New England Journal of Medicine. 368, no. 14 (2013): 1279–90.
110 Kolata G. *Mediterranean Diet Shown to Ward Off Heart Attack and Stroke*. New York Times. 25 Feb. 2013.
111 Brown D. *Mediterranean Diet Reduces Cardiovascular Risk.* Washington Post. 25 Feb. 2013.
112 *Landmark Clinical Study Reports Mediterranean Diet Supplemented with Walnuts Significantly Reduces Risk of Stroke and Cardiovascular Diseases*. California Walnut Commission. News release. 25 Feb. 2013.
113 "Доказало": *Mediterranean Diet Helps Cut Risk of Heart Attack, Stroke: Results of* PREDIMED *Study Presented*. Universitat de Barcelona. News release. 25 Feb. 2013. "Убедительно": Guasch-Ferré M. et al. *The* PREDIMED *Trial, Mediterranean Diet and Health Outcomes: How Strong Is the Evidence?* Nutrition, Metabolism and Cardiovascular Diseases. 27, no. 7 (2017): 6.
114 Carlisle J. B. *Data Fabrication and Other Reasons for Non-Random Sampling in 5087 Randomised, Controlled Trials in Anaesthetic and General Medical Journals*. Anaesthesia. 72, no. 8 (2017): 944–52.
115 Estruch R. et al. *Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts*. New England Journal of Medicine. 378, no. 25 (2018): e34.

116 Количество цитирований — по данным *Google Scholar.*
117 Belluz J. *This Mediterranean Diet Study Was Hugely Impactful. The Science Has Fallen Apart.* Vox. 13 Feb. 2019.
118 Кроме того, раз исследование завершили раньше, величина эффекта могла оказаться завышенной. См. Bassler D. et al. *Early Stopping of Randomized Clinical Trials for Overt Efficacy Is Problematic*. Journal of Clinical Epidemiology. 61, no. 3 (2008): 241–6.
119 Agarwal A., Ioannidis J. P. A. PREDIMED *Trial of Mediterranean Diet: Retracted, Republished, Still Trusted?* BMJ. (2019): 1341. Тот факт, что диета влияла только на инсульты, но не на инфаркты или смертность, был затушеван за счет использования в исследовании так называемых комбинированных конечных точек. Подобные сплавы из нескольких различных исходов часто встречаются в клинических исследованиях, поскольку увеличивают статистическую мощность. Оборотная сторона в том, что становится трудно интерпретировать конкретные эффекты изучаемого вмешательства. См. McCoy C. *Understanding the Use of Composite Endpoints in Clinical Trials*. Western Journal of Emergency Medicine. 19, no. 4 (2018): 631–4. Также см. Lim E. et al. *Composite Outcomes in Cardiovascular Research: A Survey of Randomized Trials*. Annals of Internal Medicine. 149, no. 9 (2008): 612–7. Наконец, авторы протестировали множество различных исходов, но не сделали никакой поправки на значительное количество вычисленных *p*-значений, что повышает риск ложноположительных результатов.
120 Аббревиатура OPERA означает *Oscillation Project with Emulsion-tRacking Apparatus* ("Проект по изучению нейтринных осцилляций с помощью эмульсионного трекового детектора"). Общая идея состояла в том, чтобы наблюдать, как нейтрино меняют свои свойства ("осциллируют") по мере движения от излучателя в Швейцарии до детектора в Италии. Более подробную информацию см. на сайте проекта: *http://operaweb.lngs.infn.it/spip.php?rubrique1&lang=en.*
121 Stephens R. *The Data That Threatened to Break Physics*. Nautilus. 28 Dec. 2017.
122 Adam T. et al. *Measurement of the Neutrino Velocity with the* OPERA *Detector in the CNGS Beam*. Journal of High Energy Physics. 2012, no. 10 (2012): 93.
123 CERN. OPERA *Experiment Reports Anomaly in Flight Time of Neutrinos from CERN to Gran Sasso*. 23 Sept. 2011.
124 *CERN Scientists "Break the Speed of Light"*. Daily Telegraph. 22 Sept. 2011; *The Speed of Light: Not So Fast?* ABC News. 24 Sept. 2011.
125 Вообще занижение было на 73 наносекунды, но исследователи потом обнаружили еще одну проблему, которая вызывала небольшое завышение. Совокупно две эти проблемы приводили к занижению

времени на 60 наносекунд. См. Stephens R. *The Data That Threatened to Break Physics*.

126 Grossman L. *Faster-than-Light Neutrino Result to Get Extra Checks*. New Scientist. 25 Oct. 2011.

127 Ereditato A. OPERA*: Ereditato's Point of View*. Le Scienze. 30 March 2012.

128 Palmer J. *Faster-than-Light Neutrinos Could Be down to Bad Wiring*. BBC News. 23 Feb. 2012; Grossman L., Biever C. *Was Speeding Neutrino Claim a Human Error?* New Scientist. 23 Feb. 2012.

Глава 7. **Порочные стимулы**

1 McCarthy C. *No Country for Old Men*. London: Picador, 2005. [Маккарти К. *Старикам тут не место*. М.: Азбука, 2009.]

2 Lewis S. *Cleaning Up: Inside the Wildfire Debris Removal Job That Cost Taxpayers* $ *1.3 Billion*. KQED. 19 July 2018.

3 Схожий порочный стимул связан с законом США об исчезающих видах, который поощрял землевладельцев уничтожать прекрасные места обитания редких животных, поскольку это позволяло избежать регулирования земель. Byl J. P. *Accurate Economics to Protect Endangered Species and Their Critical Habitats.* SSRN. 2018.

4 Funk C., Hefferon M. *As the Need for Highly Trained Scientists Grows, a Look at Why People Choose These Careers*. Fact Tank. 24 Oct. 2016.

5 Anderson M. S. et al. *Extending the Mertonian Norms: Scientists' Subscription to Norms of Research*. Journal of Higher Education. 81, no. 3 (2010): 366–93. Однако они не обязательно согласны с тем, что коллеги придерживаются этих норм. См. Anderson M. et al. *Normative Dissonance in Science: Results from a National Survey of U. S. Scientists*. Journal of Empirical Research on Human Research Ethics. 2, no. 4 (2007): 3–14.

6 Darwin Correspondence Project. Letter no. 5986 (6 March 1868).

7 "Четыреста тысяч новых статей": Kelly S. *The Continuing Evolution of Publishing in the Biological Sciences*. Biology Open. 7, no. 8 (2018): bio037325. "2,4 миллиона новых работ": Plume A., van Weijen D. *Publish or Perish? The Rise of the Fractional Author…* Research Trends. Sept. 2014. Согласно отчету Национального научного фонда США, в 2016 году Китай стал первой страной по количеству публикуемых научных статей, обойдя Соединенные Штаты. Tollefson J. *China Declared World's Largest Producer of Scientific Articles*. Nature. 553, no. 7689 (2018): 390.

8 Bornmann L., Mutz R. *Growth Rates of Modern Science: A Bibliometric Analysis Based on the Number of Publications and Cited References*. Journal of the Association for Information Science and Technology. 66, no. 11 (2015): 2215–22.

9 В ходе одного расследования всплыли сведения о вознаграждениях в размере 165 тысяч долларов США, что может раз в двадцать превышать годовую зарплату. В среднем суммы около 44 тысяч долларов. Все это, кстати, подчеркивает, насколько низкие у китайских ученых зарплаты, которые, по данным из статьи, представленной ниже, в среднем составляют около 8,6 тысячи долларов США в год. Иногда только часть присужденных денег становится индивидуальной премией, а остальное вкладывается в будущие исследования ученых. Реальные суммы, как и то, на что они расходуются, часто неведомы. Quan W. et al. *Publish or Impoverish: An Investigation of the Monetary Reward System of Science in China (1999–2016)*. Aslib Journal of Information Management. 69, no. 5 (2017): 486–502.

10 Abritis A. *Cash Bonuses for Peer-Reviewed Papers Go Global.* Science. 10 Aug. 2017. См. также Editorial. *Don't Pay Prizes for Published Science*. Nature. 547, no. 7662 (2017): 137.

11 Политика вроде бы действительно работает, по крайней мере в узких целях — поощрять попытки публиковаться в высокорейтинговых журналах. В одной работе было показано, что после введения этими странами денежных вознаграждений в журнал *Science* оттуда стали подавать на 46 % статей больше (существенный прирост по сравнению с другими мерами поощрения) — хотя реальная доля принятых из них к публикации уменьшилась. То есть ученые стали чаще поглядывать в сторону высокорейтинговых журналов, но без особого успеха. См. Franzoni C. et al. *Changing Incentives to Publish*. Science. 333, no. 6043 (2011): 702–3. Иронично: один из авторов этой статьи, которая и сама была опубликована в *Science*, получил за нее денежное вознаграждение в размере трех с половиной тысяч долларов — правда, он отдал их на благотворительность. См. Abritis A. *Cash Bonuses.*

12 *www.ref.ac.uk.* В других странах шло обсуждение — и в итоге решено было не использовать подобную практику: Sivertsen G. *Why Has No Other European Country Adopted the Research Excellence Framework?* LSE Impact of Social Sciences. 18 Jan. 2018.

13 Попытку (не вполне убедительную) проследить происхождение фразы “Публикуйся или погибни” см. здесь: Garfield E. *What is the Primordial Reference for the Phrase “Publish or Perish”?* Scientist. 10, no. 2 (1996): 11.

14 Link A. N. et al. *A Time Allocation Study of University Faculty*. Economics of Education Review. 27, no. 4 (2008): 363–74.

15 Цитирую по Библии короля Якова. Впервые в контексте науки эта цитата прозвучала из уст Роберта Мертона, сформулировавшего мертоновские нормы. Merton R. K. *The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered*. Science. 159, no. 3810 (1968): 56–63.

16 Bol T. et al. *The Matthew Effect in Science Funding*. Proceedings of the National Academy of Sciences. 115, no. 19 (2018): 4887–90.

17 Это подкрепляется двумя линиями доказательств. Во-первых, подавляющее большинство тех, кто получает ученую степень, не остаются в науке навсегда. В исследовании, проведенном в 2010 году для Королевского общества, выяснилось, что 53 % получивших ученую степень сразу же ушли из науки, а затем к ним присоединились еще 26,5 %, оставив науку после завершения начального этапа карьеры. 17 % ушли, чтобы заниматься исследованиями вне академических кругов, например в промышленности или правительстве. В целом только 3,5 % остались в науке навсегда (а 0,45 % дошли до полной профессуры). *The Scientific Century: Securing Our Future Prosperity.* The Royal Society. 2010. Вторая линия доказательств — результаты опросов ученых. В одном таком опросе, проведенном фондом *Wellcome Trust* в начале 2020 года, не менее 78 % респондентов согласились, что высочайшая конкуренция в науке "создала жесткие и агрессивные условия". *What Researchers Think about the Culture They Work In*. Wellcome Trust. 2020.

18 Brischoux F., Angelier F. *Academia's Never-Ending Selection for Productivity*. Scientometrics. 103, no. 1 (2015): 333–6. В других исследованиях результаты сходные. Дополнительное свидетельство, касающееся канадского рынка труда в области психологии, см. тут: Pennycook G., Thompson V. A. *An Analysis of the Canadian Cognitive Psychology Job Market (2006–2016).* Canadian Journal of Experimental Psychology. 72, no. 2 (2018): 71–80.

19 Cyranoski D. et al. *Education: The PhD Factory*. Nature. 472, no. 7343 (2011): 276–9. Возможно, вы удивляетесь, почему, если так много неприкаянных обладателей ученой степени слоняются без исследовательской работы, часто говорят о кризисе — когда в область науки, технологий, инженерии и математики идет трудиться меньше людей, чем того требует современная индустриальная экономика. В одном обзоре было показано, что на самом деле верно и то и другое: слишком много обладателей ученой степени ищут работу в университетском секторе, но недостаточно — в правительстве и промышленности. Xue Y., Larson R. *STEM Crisis or STEM Surplus? Yes and Yes*. Monthly Labor Review. 26 May 2015.

20 Heitz R. P. *The Speed-Accuracy Tradeoff: History, Physiology, Methodology, and Behavior.* Frontiers in Neuroscience. 8 (2014): 150.

21 Это установлено математически: Heesen R. *Why the Reward Structure of Science Makes Reproducibility Problems Inevitable*. Journal of Philosophy. 115, no. 12 (2018): 661–74. Еще см. Sarewitz D. *The Pressure to Publish Pushes Down Quality*. Nature. 533, no. 7602 (2016): 147.

22 При всем при том я должен отметить работу, вроде бы противоречащую моим аргументам: Fanelli D. et al. *Misconduct Policies, Aca-*

demic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. PLOS ONE. 10, no. 6 (2015): e0127556. В 2015 году ее авторы подсчитали все научные статьи, которые были публично исправлены или отозваны в 2010-м и 2011-м, то есть, по сути, все случаи за эти годы, когда приходилось вносить исправления в научную летопись, и для каждого ученого соотнесли количество случаев, когда ему приходилось делать исправления, с другими его характеристиками, такими как общее число публикаций и страна, где он работал. Общий вывод был следующим: в странах, где научным работникам платят за статьи, и в тех, где не проводится надлежащая политика по борьбе с недобросовестной научной практикой, чаще происходит отзыв статей. Пока с моим тезисом все согласуется. Однако еще авторы того исследования обнаружили, что на счету ученых, публикующих больше статей в год, в целом отзывов статей немножко меньше, — они интерпретировали этот вывод как противоречащий идее о том, будто культура “публикуйся или погибни” вызывает больше нарушений научных норм. Правда, статьи отзывают довольно редко, ведь это крайняя мера: отзыв статьи подразумевает ее выкорчевывание из научной литературы и часто происходит из-за серьезных нарушений типа мошенничества. Поэтому, на мой взгляд, это исследование 2015 года не может приводиться в защиту принципа “публикуйся или погибни”, ведь никак не учитывалось качество статей. Авторы также выяснили, что ученые, которые публикуются регулярнее, как правило, публикуют и больше исправлений. По их словам, это хорошо, ведь исправления “не несут на себе клейма”. Я нахожу этот аргумент странным: по моему опыту, исправления не только уж точно несут на себе клеймо, но еще и по определению подтверждают ошибки: исправляются изъяны, которых быть вообще не должно было.

23 Quan W. et al. *Publish or Impoverish.*

24 Первое упоминание термина “нарезка” в научном контексте, которое мне удалось найти: Maddox J. *Is the Salami Sliced Too Thinly?* Nature. 342, no. 6251 (1989): 733. В статье говорится, что эта метафора уже какое-то время использовалась и ранее. Кроме того, термин употреблялся и в других контекстах, например, в случаях, когда работники крали крупные суммы у своего работодателя, долгое время систематически присваивая себе понемножку.

25 Я отказываюсь приводить здесь все эти статьи, поскольку авторы того и добиваются. Вот одна из них: Chen X. et al. *A Novel Relationship for Schizophrenia, Bipolar and Major Depressive Disorder Part 5: A Hint from Chromosome 5 High Density Association Screen*. American Journal of Translational Research. 9, no. 5 (2017): 2473–91.

26 Кстати, ни в той ни в другой не было найдено никаких различий — так что как минимум эти статьи сообщали об отрицательных ре-

зультатах. Spielmans G. I. et al. *A Case Study of Salami Slicing: Pooled Analyses of Duloxetine for Depression*. Psychotherapy and Psychosomatics. 79, no. 2 (2010): 97–106.

27 Другой пример, уже про нейролептики, см. здесь: Spielmans G. I. et al. *“Salami Slicing” in Pooled Analyses of Second-Generation Antipsychotics for the Treatment of Depression*. Psychotherapy and Psychosomatics. 86, no. 3 (2017): 171–2.

28 Конечно, важная мысль этой книги в том, что со стандартами рецензирования и другими правилами, принятыми в научных журналах, все весьма плохо — иначе и быть не может, учитывая количество низкокачественных статей, которые принимаются и публикуются. Но хищнические журналы даже и не пытаются ничего делать.

29 Есть похожая мошенническая индустрия фиктивных научных *конференций*, куда ученые приглашаются в постоянных спам-рассылках. Вот хорошие обзоры: McCrostie J. *“Predatory Conferences” Stalk Japan’s Groves of Academia*. Japan Times. 11 May 2016; Stoye E. *Predatory Conference Scammers Are Getting Smarter.* Chemistry World. 6 Aug. 2018.

30 Джеффри Белл, библиотекарь из Колорадского университета в Денвере, проводит единоличную кампанию против хищнических журналов. См. Beall J. *What I Learned from Predatory Publishers*. Biochemia Medica. 27, no. 2 (2017): 273–8. Составленный им перечень подозрительных издательств в конце концов исчез из интернета (*https://retractionwatch.com/2017/01/17/bealls-list-potential-predatory-publishers-go-dark)*, но более свежую версию с длиннющим списком “вероятных хищнических журналов” можно найти здесь: *https://beallslist.net/standalone-journals.* Также см. Bolshete P. *Analysis of Thirteen Predatory Publishers: A Trap for Eager-to-Publish Researchers*. Current Medical Research and Opinion. 34, no. 1 (2018): 157–62; Grudniewicz A. et al. *Predatory Journals: No Definition, No Defence*. Nature. 576, no. 7786 (2019): 210–2.

31 Увы, статья так и не была официально опубликована, поскольку Вэмплю отказался платить журналу сто пятьдесят долларов за публикацию (подробнее см. Stromberg J. *“Get Me Off Your Fucking Mailing List” Is an Actual Science Paper Accepted by a Journal*. Vox. 21 Nov. 2014). Рукопись, исходно созданная несколькими годами ранее в сходных целях специалистами по компьютерным наукам Дэвидом Мэзиерсом и Эдди Колером, во всей своей блистательной полноте располагается здесь: *www.scs.stanford.edu/~dm/home/papers/remove.pdf.*

32 Цитируется по: Oransky I. *South Korean Plant Compound Researcher Faked Email Addresses So He Could Review His Own Studies*. Retraction Watch. 24 Aug. 2012.

33 Подробное обсуждение одного случая и покаяние допустивших его редакторов см. здесь: Cohen A. et al. *Organised Crime against the Academic Peer Review System*. British Journal of Clinical Pharmacology. 81, no. 6 (2016): 1012–7.

34 McCook A. *A New Record: Major Publisher Retracting More than 100 Studies from Cancer Journal over Fake Peer Reviews*. Retraction Watch. 20 April 2017.

35 Lariviere V. et al. *The Decline in the Concentration of Citations, 1900–2007*. Journal of the American Society for Information Science and Technology. 60, no. 4 (2009): 858–62.

36 Там же. По крайней мере, в естественных науках дела не так плохи, как в гуманитарных: менее 20 % статей по гуманитарным дисциплинам цитируются кем-то в течение первых пяти лет. Разумеется, в естественных и гуманитарных науках приняты разные нормы цитирования — в последних, например, больше ссылаются на книги, чем на статьи. И все же любопытно, большой ли вклад в науку внесли эти 80 % никем не процитированных статей…

37 Hirsch J. E. *An Index to Quantify an Individual's Scientific Research Output*. Proceedings of the National Academy of Sciences. 102, no. 46 (2005): 16569–72.

38 Было бы упущением не сказать, что *Google Scholar* скорее завышает индекс Хирша, поскольку критерии того, что считать за цитирование, там не слишком строгие. Существуют и другие способы вычислить индекс Хирша, например сайт *Web of Science* — критерии там построже, и потому он дает более низкие цифры.

39 Duyx B. et al. *Scientific Citations Favor Positive Results: A Systematic Review and Meta-Analysis*. Journal of Clinical Epidemiology. 88 (2017): 92–101; также см. Leimu R., Koricheva J. *What Determines the Citation Frequency of Ecological Papers?* Trends in Ecology & Evolution. 20, no. 1 (2005): 28–32.

40 Aksnes D. W. *A Macro Study of Self-Citation*. Scientometrics. 56, no. 2 (2003): 235–46. Самоцитирование прекрасно работает в паре с нарезкой: если ваши нарезанные на кусочки статьи все ссылаются друг на друга, число цитирований у вас получится гораздо больше, даже если никто, кроме вас, на самом деле вашу работу не читал и на нее не ссылался. Велик соблазн сказать, что нарезанные кусочки накладываются один на другой и становятся этаким закольцованным веером, но, пожалуй, это заводит метафору слишком далеко.

41 О некоторых крайних формах см. Ioannidis J. P. *A Generalized View of Self-Citation: Direct, Co-Author, Collaborative, and Coercive Induced Self-Citation*. Journal of Psychosomatic Research. 78, no. 1 (2015): 7–11.

42 Flaherty C. *Revolt Over an Editor*. Inside Higher Ed. 30 April 2018. Похожая история, только связанная с исследованиями расстройств

аутистического спектра: Etchells P., Chambers C. *The Games We Play: A Troubling Dark Side in Academic Publishing.* The Guardian. 12 March 2015.

43 Fried E. *7 Sternberg Papers: 351 References, 161 Self-Citations.* Eiko Fried. 29 March 2018.

44 Thombs B. D. et al. *Potentially Coercive Self-Citation by Peer Reviewers: A Cross-Sectional Study*. Journal of Psychosomatic Research. 78, no. 1 (2015): 1–6.

45 Довольно иронично, что Стернберг пытался обмануть систему именно этим способом, поскольку в 2017 году в своей книге с советами для ученых он написал следующее: “Самоплагиат возникает, когда вы цитируете свою собственную работу неправильным образом. <… > В крайних случаях кто-то может попытаться опубликовать одну и ту же статью дважды, никак не отметив, что эта работа уже была опубликована ранее”. Sternberg R. J., Sternberg K. *The Psychologist's Companion: A Guide to Professional Success for Students, Teachers, and Researchers*. Cambridge: CUP, 2016.

46 Исходная статья (где текст был опубликован впервые): Sternberg R. J. *WICS: A New Model for Cognitive Education.* Journal of Cognitive Education and Psychology. 9, no. 1 (2010): 36–47. Сообщение об отзыве статьи как “дублирующей публикации”: Editorial. *Retraction Notice for “WICS: A New Model for School Psychology” by Robert J. Sternberg.* School Psychology International. 39, no. 3 (2018): 329. См. также Brown N. J. L. *Some Instances of Apparent Duplicate Publication by Dr. Robert J. Sternberg*. Nick Brown's Blog. 25 April 2018.

47 Bretag T., Carapiet S. *A Preliminary Study to Identify the Extent of Self-Plagiarism in Australian Academic Research*. Plagiary: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification. 2, no. 5 (2007): 1–12.

48 Archambault É., Lariviere V. *History of the Journal Impact Factor: Contingencies and Consequences*. Scientometrics. 79, no. 3 (2009): 635–49.

49 Технически, что, кстати, немного путает, импакт-фактор — это среднее за конкретный год количество ссылок на статьи, опубликованные в журнале за предыдущие два года. Понятно, что для расчета этого показателя должен пройти целый год, поэтому значение всегда несколько неактуальное. Например, в 2020 году импакт-фактор журнала будет представлять собой среднее число ссылок за 2019-й на статьи, опубликованные там в 2017 и 2018 годах.

50 Вдруг вам интересно: у журнала *American Journal of Potato Research* импакт-фактор на момент написания этих строк равен 1,095.

51 Lariviere V. et al. *A Simple Proposal for the Publication of Journal Citation Distributions*. BioRxiv (5 July 2016); также см. Lariviere V., Sugimoto C. R. *The Journal Impact Factor: A Brief History, Critique,*

and Discussion of Adverse Effects // Glänzel W. et al. (eds.) *Springer Handbook of Science and Technology Indicators*. Cham: Springer International Publishing, 2019. Аргументы в защиту импакт-фактора см. тут: Bornmann L., Pudovkin A. I. *The Journal Impact Factor Should Not Be Discarded.* Journal of Korean Medical Science. 32, no. 2 (2017): 180–2.

52 Monastersky R. *The Number That's Devouring Science*. Chronicle of Higher Education. 14 Oct. 2005.

53 Wilhite A. W., Fong E. A. *Coercive Citation in Academic Publishing*. Science. 335, no. 6068 (2012): 542–3.

54 Davis P. *The Emergence of a Citation Cartel*. Scholarly Kitchen. 10 April 2012.

55 Jump P. *Journal Citation Cartels on the Rise.* Times Higher Education. 21 June 2013. Нашелся для картелей, похоже, и собственный Элиот Несс: в 2017 году был разработан алгоритм, который при вводе данных о цитированиях выделяет группы авторов, ссылающихся друг на друга слишком часто. Fister I. Jr. et al. *Toward the Discovery of Citation Cartels in Citation Networks.* Frontiers in Physics. 4:49 (2016).

56 Goodhart C. *Monetary Relationships: A View from Threadneedle Street // Papers in Monetary Economics.* Vol. I. Reserve Bank of Australia, 1975. Конкретная формулировка взята отсюда: Strathearn M. *Improving Ratings: Audit in the British University System.* European Review. 5, no. 3 (1997): 305–21.

57 Smaldino P. E., McElreath R. *The Natural Selection of Bad Science*. Royal Society Open Science. 3, no. 9 (2016): 160384.

58 Higginson A. D., Munafo M. R. *Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions*. PLOS Biology. 14, no. 11 (2016): 6.

59 Grimes D. R. et al. *Modelling Science Trustworthiness under Publish or Perish Pressure*. Royal Society Open Science. 5, no. 1 (2018).

60 У этого художника есть и другая версия картины, с другим алхимиком. Та размещена в Национальной галерее искусства в Вашингтоне.

61 Howes A. *Age of Invention: When Alchemy Works*. Age of Invention. 6 Oct. 2019; Conniff R. *Alchemy May Not Have Been the Pseudoscience We All Thought It Was*. Smithsonian Magazine. Feb. 2014.

62 Существенная разница между алхимиками и современными учеными заключается в том, что для современных ученых золото реально. Как мы видели в шестой главе, ученые, которые охотно раздувают и упрощают до абсурда свои результаты, получают огромные денежные вознаграждения благодаря книгам-бестселлерам и масштабным лекционным турам.

63 Edwards M. A., Roy S. *Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and*

Hypercompetition. Environmental Engineering Science. 34, no. 1 (2017): 51–61.
64 Yarkoni T. *No, it's not The Incentives — it's you*. [Citation Needed]. 2 Oct. 2018.
65 См. Edwards M. A., Roy S. *Academic Research in the 21st Century*. Рис. 1.
66 Летом 2020 года мы с карикатуристом Заком Винерсмитом (автором блестящего веб-комикса *Saturday Morning Breakfast Cereal*) создали комикс по мотивам этой главы: *www.smbc-comics.com/comic/science-fictions.*

Глава 8. **Исцеление науки**

1 Nielsen M. *The Future of Science*. 17 July 2008. *http://michaelnielsen.org/blog/the-future-of-science-2/*
2 De Vries Y. A. et al. *The Cumulative Effect of Reporting and Citation Biases on the Apparent Efficacy of Treatments: The Case of Depression*. Psychological Medicine. 48, no. 15 (2018): 2453–5.
3 Там же.
4 А именно, они обнаружили примеры подачи под нужным углом и предвзятости цитирования; другие проблемы выявить было сложнее, поскольку испытания методов психотерапии, в отличие от испытаний новых лекарств, не требуется предварительно регистрировать.
5 В некоторых странах так происходит в том числе из-за законов о конфиденциальности. Seife C. *Research Misconduct Identified by the US Food and Drug Administration: Out of Sight, Out of Mind, Out of the Peer-Reviewed Literature*. JAMA Internal Medicine. 175, no. 4 (2015): 567–77. Еще см. Robinson M. *Canadian Researchers Who Commit Scientific Fraud Are Protected by Privacy Laws*. The Star. 12 July 2016.
6 Oransky I., Marcus A. *Governments Routinely Cover up Scientific Misdeeds. Let's End That*. STAT News. 15 Dec. 2015.
7 Hou C.-Y. *Sweden Passes Law For National Research Misconduct Agency*. Scientist. 10 July 2019.
8 Oksvold M. P. *Incidence of Data Duplications in a Randomly Selected Pool of Life Science Publications*. Science and Engineering Ethics. 22, no. 2 (2016): 487–96; Bucci M. E. et al. *Automatic Detection of Image Manipulations in the Biomedical Literature*. Cell Death & Disease. 9, no. 3 (2018): 400. Кстати, в этой последней статье отмечается, что алгоритм искусственного интеллекта выявил “пугающее” число дублирований изображений в литературе по клеточной биологии.
9 Технологии также могут помочь ученым яснее и четче видеть научную литературу. В третьей главе мы обсудили проблему статей-зомби, на которые даже после их отзыва продолжают ссылаться.

Есть автоматизированное решение этой проблемы: компания *Zotero*, выпускающая бесплатное программное обеспечение для работы со ссылками (многие ученые используют его для создания библиографического списка при написании каждой статьи, что уже значительно экономит время и предотвращает ошибки; я использовал его при создании этой книги), недавно объявила о партнерстве с *Retraction Watch*. Благодаря этому сотрудничеству программа будет отмечать любые изъятые из научной летописи публикации, которые вы собираетесь процитировать, информируя о том, что “эта работа была отозвана”. Stillman D. *Retracted Item Notifications with Retraction Watch Integration*. Zotero. 14 June 2019.

10 Предполагают даже, что сложные алгоритмы могли бы сканировать десятки тысяч статей и находить все особенности (возможно, неочевидные для человека), которые позволяют предсказать, будут ли описанные там результаты воспроизводиться. Rogers A. *Darpa Wants to Solve Science's Reproducibility Crisis With AI*. Wired. 15 Feb. 2019.

11 Одна из таких программ — *RMarkdown* (*https://rmarkdown.rstudio.com*).

12 Еще более радикальное предложение внес психолог Джефф Роудер — об “открытых с рождения данных”, когда данные каждого нового участника эксперимента в конце каждого экспериментального дня автоматически загружаются онлайн. Rouder J. N. *The What, Why, and How of Born-Open Data*. Behavior Research Methods. 48, no. 3 (2016): 1062–9.

13 Ziemann M. et al. *Gene Name Errors are Widespread in the Scientific Literature*. Genome Biology. 17, no. 1 (2016): 177.

14 *Ottoline Leyser on How Plants Decide What to Do*. The Life Scientific. BBC Radio 4. 16 May 2017.

15 Nosek B. A. et al. *Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability*. Perspectives on Psychological Science. 7, no. 6 (2012): 615–31.

16 Архив статей журнала *Journal of Negative Results in Biomedicine*: *https://jnrbm.biomedcentral.com/articles.* Еще есть *Journal of Negative Results: Ecology and Evolutionary Biology* — издание, выглядящее менее профессионально, куда подавалось на рассмотрение ничтожное количество работ (*www.jnr-eeb.org/index.php/jnr).* В 2014 и 2015 годах там не было опубликовано ни единой статьи, в 2016-м вышло две, затем в 2018-м еще одна, а с тех пор о нем ничего не слышно [потом в 2020 году вышло две статьи, в 2021-м и 2022-м — по одной, в 2023-м — пока две *(прим. перев.)*]. Едва ли можно подобрать лучшую иллюстрацию того, насколько непопулярным считается опубликовать свою статью в журнале, принимающем только работы с отрицательными результатами.

17 PLOS означает *Public Library of Science* ("публичная научная библиотека"). Стратегия журнала PLOS ONE, публикующего все, что прошло этап рецензирования, независимо от "интересности", стала поразительно успешной, превратив его в самое плодовитое (по числу публикуемых работ) издание мира. Так было до 2017 года, когда его обошел мегажурнал *Scientific Reports* с такой же публикационной стратегией. Davis P. *Scientific Reports Overtakes PLoS ONE As Largest Megajournal*. 6 April 2017. Журнал *Scientific Reports* издается *Nature Publishing Group*, компанией, выпускающей и самый топовый журнал в мире — *Nature*, так что популярность *Scientific Reports* отчасти может объясняться неким эффектом ореола. Вот еще пара примеров журналов со стратегией "принимать к публикации все высококачественное, независимо от предполагаемого воздействия": *PeerJ* (*https://peerj.com*) и *Royal Society Open Science* (*https://royalsocietypublishing.org/journal/rsos*).

18 *www.apa.org/pubs/journals/psp/?tab=4*

19 Srivastava S. *A Pottery Barn Rule for Scientific Journals*. The Hardest Science. 27 Sept. 2012. Это именуют "правилом гончарной мастерской" *(Pottery Barn Rule)*, по названию сети американских магазинов *Pottery Barn*, где оно якобы применяется. Однако это что-то вроде городской легенды, поскольку в магазинах *Pottery Barn* такого правила нет. См. Grant D. *You Break It, You Buy It? Not According to the Law*. Crafts Report. April 2005.

20 Nosek B. A. et al. *Promoting an Open Research Culture*. Science. 348, no. 6242 (2015): 1422–5.

21 De Vrieze J. *"Replication Grants" Will Allow Researchers to Repeat Nine Influential Studies That Still Raise Questions*. Science. 11 July 2017.

22 Приведу лишь два хорошо изученных примера, относящихся к середине и концу XX века: Bakan D. *The Test of Significance in Psychological Research*. Psychological Bulletin. 66, no. 6 (1966): 423–37; Cohen J. *The Earth Is Round (p < .05)*. American Psychologist. 49, no. 12 (1994): 997–1003.

23 Ziliak S. T., McCloskey D. N. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives, Economics, Cognition, and Society*. Ann Arbor: University of Michigan Press, 2008.

24 Amrhein V. et al. *Scientists Rise up against Statistical Significance.* Nature. 567, no. 7748 (2019): 305–7.

25 Подробнее эта точка зрения изложена здесь: Cumming G. *The New Statistics: Why and How*. Psychological Science. 25, no. 1 (2014): 7–29; Halsey L. G. *The Reign of the p-Value Is Over: What Alternative Analyses Could We Employ to Fill the Power Vacuum?* Biology Letters. 15, no. 5 (2019): 20190174.

26 Если известны величина эффекта и p-значение, вы можете вычислить доверительный интервал, и наоборот. См. Altman D. G.,

Bland J. M. *How to Obtain the Confidence Interval from a P Value*. BMJ. 343 (2011): d2090; Altman D. G., Bland J. M. *How to Obtain the P Value from a Confidence Interval*. BMJ. 343 (2011): d2304.

27 Ioannidis J. P. A. *The Importance of Predefined Rules and Prespecified Statistical Analyses: Do Not Abandon Significance.* JAMA. 321, no. 21 (2019): 2067.

28 Цитируется по: Gelman A. *"Retire Statistical Significance": The Discussion*. Statistical Modeling, Causal Inference, and Social Science. 20 March 2019.

29 В этом смысле байесовская априорность — это математическое воплощение знаменитой фразы Карла Сагана "Экстраординарные утверждения требуют экстраординарных доказательств".

30 К слову, более распространенная статистическая традиция, где присутствуют *p*-значения, называется *частотной* статистикой (частотным подходом). Потому что тех, кто использует *p*-значения, по сути, интересуют *частоты* — главным образом частота, с которой вы получали бы результаты с *p*-значениями ниже 0,05, если бы проверяемая вами гипотеза была неверна, а исследование проводилось бы бесконечное число раз.

31 Полезный аннотированный список литературы, который служит введением в байесовскую статистику, см. здесь: Etz A. et al. *How to Become a Bayesian in Eight Easy Steps: An Annotated Reading List*. Psychonomic Bulletin & Review. 25, no. 1 (2018): 219–34. Еще см. McElreath R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton: CRC Press / Taylor & Francis Group, 2016.

32 Доводы в защиту (зачастую весьма квалифицированную) затравленных *p*-значений см. тут: Savalei V., Dunn E. *Is the Call to Abandon P-Values the Red Herring of the Replicability Crisis?* Frontiers in Psychology. 6:245 (2015); Murtaugh P. A. *In Defense of P Values.* Ecology. 95, no. 3 (2014): 611–7; Senn S. *Two Cheers for P-Values?* Journal of Epidemiology and Biostatistics. 6, no. 2 (2001): 193–204.

33 Benjamin D. J. et al. *Redefine Statistical Significance*. Nature Human Behaviour. 2, no. 1 (2018): 6–10. Впрочем, см. и ответ: Lakens D. et al. *Justify Your Alpha*. Nature Human Behaviour. 2, no. 3 (2018): 168–71. Есть также новые статистические методы, в которых все еще используются *p*-значения, но отброшен наивный сценарий, когда результаты всегда сравниваются с нулем (или нулевой разницей). См. Lakens D. et al. *Equivalence Testing for Psychological Research: A Tutorial.* Advances in Methods and Practices in Psychological Science. 1, no. 2 (2018): 259–69.

34 Либо же исследователи могли бы проводить анализ данных самостоятельно, но при этом позволять независимым статистикам его интерпретировать. Boutron I., Ravaud P. *Misrepresentation and*

Distortion of Research in Biomedical Literature. Proceedings of the National Academy of Sciences. 115, no. 11 (2018): 2613–9.

35 Или когда статистик считает, что плох сам дизайн исследования. Известно такое высказывание Рональда Фишера 1938 года: "Зачастую консультироваться со статистиком после завершения эксперимента — значит просто просить его провести посмертное обследование. Возможно, он сумеет сказать, от чего эксперимент умер". Fisher R. *Presidential Address [to the First Indian Statistical Congress]*. Sankhyā: The Indian Journal of Statistics (1933–1960). 4, no. 1 (1938): 14–7.

36 "Анализ кривой спецификаций": Simonsohn U. et al. *Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications*. SSRN. 2015. "Анализ вибраций эффектов": Patel C. J. et al. *Assessment of Vibration of Effects Due to Model Specification Can Demonstrate the Instability of Observational Associations*. Journal of Clinical Epidemiology. 68, no. 9 (2015): 1046–58. "Анализ мультивселенной": Steegen S. et al. *Increasing Transparency Through a Multiverse Analysis.* Perspectives on Psychological Science. 11, no. 5 (2016): 702–12.

37 Orben A., Przybylski A. K. *The Association between Adolescent Well-Being and Digital Technology Use*. Nature Human Behaviour. 3, no. 2 (2019): 173–82. Признание: Орбен и Пшибыльски — мои друзья и коллеги.

38 Blanchard S. *Smartphones and Tablets Are Causing Mental Health Problems in Children as Young as TWO by Crushing Their Curiosity and Making Them Anxious*. MailOnline. 2 Nov. 2018.

39 См., например: Twenge J. M. *Have Smartphones Destroyed a Generation?* Atlantic. Sept. 2017; Twenge J. M. *IGEN: Why Today's Super-Connected Kids Are Growing up Less Rebellious, More Tolerant, Less Happy — and Completely Unprepared for Adulthood (and What This Means for the Rest of Us)*. New York: Atria Books, 2017.

40 "Видеоигровое расстройство": *www.who.int/features/qa/gaming-disorder/en/;* Van Rooij A. J. et al. *A Weak Scientific Basis for Gaming Disorder: Let Us Err on the Side of Caution*. Journal of Behavioral Addictions. 7, no. 1 (2018): 1–9. "Зависимость от онлайн-порнографии": De Alarcón R. et al. *Online Porn Addiction: What We Know and What We Don't — A Systematic Review.* Journal of Clinical Medicine. 8, no. 1 (2019): 91. "Зависимость от айфона": Spicer A. *The iPhone Is the Crack Cocaine of Technology. Don't Celebrate Its Birthday.* The Guardian. 29 June 2017. "Список можно продолжить": Snowdon C. *Evidence-Based Puritanism*. Velvet Glove, Iron Fist. 10 Jan. 2019.

41 Эти же ученые провели и другие исследования, применяя анализ мультивселенной к различным наборам данных, и вывод в целом получался такой же: паника по поводу "экранного времени" раздута. См., например: Orben A., Przybylski A. K. *Screens, Teens,*

and Psychological Well-Being: Evidence from Three Time-Use-Diary Studies. Psychological Science. 30, no. 5 (2019): 682–96; Orben A. et al. *Social Media's Enduring Effect on Adolescent Life Satisfaction.* Proceedings of the National Academy of Sciences. 116, no. 21 (2019): 10226–8. Это не значит, что все действия в интернете безвредны или что экранное время не вызывает у некоторых детей проблем. Однако в среднем влияние гораздо меньше, чем уверяют вас средства массовой информации, сеющие панику.

42 Хронологию можно посмотреть на сайте *ClinicalTrials.gov* по этой ссылке: *https://clinicaltrials.gov/ct2/about-site/history.* См. также Todd J. L. et al. *Using ClinicalTrials.Gov to Understand the State of Clinical Research in Pulmonary, Critical Care, and Sleep Medicine*. Annals of the American Thoracic Society. 10, no. 5 (2013): 411–7.

43 Scott S. *Pre-Registration Would Put Science in Chains*. Times Higher Education. 25 July 2013.

44 Wagenmakers E.-J. et al. *An Agenda for Purely Confirmatory Research*. Perspectives on Psychological Science. 7, no. 6 (2012): 632–8.

45 Kaplan R. M., Irvin V. L. *Likelihood of Null Effects of Large NHLBI Clinical Trials has Increased Over Time*. PLOS ONE. 10, no. 8 (2015): e0132382.

46 Следует отметить, что у нас нет еще ответа на главный вопрос: чаще ли результаты предварительно зарегистрированных исследований будут воспроизводиться в будущем? Только время и накапливающиеся метанаучные данные покажут это.

47 Anderson K. *Why Is ClinicalTrials.Gov Still Struggling?* Scholarly Kitchen. 15 March 2016; Anderson M. *Compliance with Results Reporting at ClinicalTrials.Gov*. New England Journal of Medicine. 372, no. 11 (2015): 1031–9; Chen R. et al. *Publication and Reporting of Clinical Trial Results: Cross Sectional Analysis across Academic Medical Centers*. BMJ. (2016): i637; Goldacre B. et al. *Compliance with Requirement to Report Results on the EU Clinical Trials Register: Cohort Study and Web Resource*. BMJ. (2018): k3218. Еще есть некоторые предварительные свидетельства из области психологии, показывающие, что мечта о заранее зарегистрированном анализе не всегда отвечает реальности: Claesen A. et al. *Preregistration: Comparing Dream to Reality*. Preprint, PsyArXiv (9 May 2019).

48 "Одно расследование, проведенное журналом *Science*": Piller C. *FDA and NIH Let Clinical Trial Sponsors Keep Results Secret and Break the Law*. Science. 13 Jan. 2020. "Нарушители": Piller C. *Clinical Scofflaws?* Science. 13 Jan. 2020.

49 Turner S. D. et al. *Publication Rate for Funded Studies from a Major UK Health Research Funder: A Cohort Study.* BMJ Open. 3, no. 5 (2013): e002521; Chinnery F. et al. *Time to Publication for NIHR HTA Programme-Funded Research: A Cohort Study.* BMJ Open.

3, no. 11 (2013): e004121. Еще см. Glasziou P., Chalmers I. *Funders and Regulators Are More Important than Journals in Fixing the Waste in Research.* The BMJ Opinion. 6 Sept. 2017.

50 Chambers C. *Registered Reports: A New Publishing Initiative at Cortex*. Cortex. 49, no. 3 (2013): 609–10. На момент написания этих строк 225 журналов [уже более трехсот *(прим. перев.)*] предлагают такую опцию — капля в море, но она растет (*https://cos.io/rr).* Также см. Chambers C. *The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice*. Princeton: Princeton University Press, 2017.

51 Начальный этап работы с зарегистрированными отчетами показал почти такую же поразительную картину, какую демонстрировал рисунок 4: 61 % отрицательных результатов для зарегистрированных психологических и биомедицинских исследований по сравнению с примерно 10 % для обычных, незарегистрированных исследований. Allen C., Mehler D. M. A. *Open Science Challenges, Benefits and Tips in Early Career and Beyond*. PLOS Biology. 17, no. 5 (2019): e3000246 (см. в этой статье рис. 1); также см. публикацию с почти идентичными результатами для психологии: Scheel A. M. et al. *An Excess of Positive Results: Comparing the Standard Psychology Literature with Registered Reports.* Preprint, PsyArXiv (5 Feb. 2020). Опять-таки существуют и иные объяснения такой разницы. Если виды исследований, где ученые склонны использовать зарегистрированные отчеты, отличаются от тех, где такие отчеты распространены меньше (например, ученые могут чаще прибегать к зарегистрированным отчетам в тех случаях, когда крайне скептически относятся к результатам, а следовательно, вероятность того, что эти результаты верны, может быть ниже), логично ожидать различия по количеству положительных и отрицательных результатов, которое никак не связано с зарегистрированным отчетом *как таковым*.

52 Вообще говоря, можно даже считать учреждение в XVII веке Королевским обществом научного журнала одним из первых шагов к открытости науки и превращению ее в нечто большее, чем просто частное занятие отдельных людей. См. David P. A. *The Historical Origins of "Open Science": An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution*. Capitalism and Society. 3, no. 2 (2008).

53 Полезный обзор: Munafo M. R. et al. *A Manifesto for Reproducible Science.* Nature Human Behaviour. 1, no. 1 (2017): 0021.

54 *Open Science Framework* (*https://osf.io)* — онлайн-репозиторий, где можно хранить данные, а также материалы предварительной регистрации с отметками времени, препринты статей и многое другое.

55 Это реализовано в журнале по общей биологии *eLife (https://elifesciences.org):* когда вы смотрите любую из опублико-

ванных там статей, с ее веб-страницы можно по ссылкам перейти к предыдущим версиям и рецензиям.

56 Но такое случается. На самом деле, пока я пишу эти строки, как раз такой случай, по-видимому, происходит в области поведенческой экологии. См. Viglione G. *"Avalanche" of Spider-Paper Retractions Shakes Behavioural-Ecology Community*. Nature. 578, no. 7794 (2020): 199–200.

57 Не так-то это просто — разместить где-нибудь в интернете данные. Имеет смысл организовать все так, чтобы была запись, у кого есть доступ к данным, то есть чтобы существовал своего рода охранник, к которому ученые могут обращаться с запросами. Принцип тот же, что и с регистрацией: это преграда для исследователей, которым может захотеться скачать находящийся в свободном доступе набор данных, прибегнуть к *p*-хакингу, чтобы получить значимые результаты, а затем опубликовать статью, сделав вид, будто они с самого начала предполагали получить такие результаты. Правда, охрана, наряду с репозиториями, где хранятся данные, требует финансирования. Это одна из причин, почему ученые редко делали так в прошлом, и одна из проблем, с решением которой могли бы помочь финансирующие организации.

58 Herfst S. et al. *Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets*. Science. 336, no. 6088 (2012): 1534–41.

59 Committee on Science, Technology, and Law et al. *Perspectives on Research with H5N1 Avian Influenza: Scientific Inquiry, Communication, Controversy: Summary of a Workshop*. Washington, D. C.: National Academies Press, 2013.

60 Stokols D. et al. *The Science of Team Science*. American Journal of Preventive Medicine. 35, no. 2 (2008): S77–89.

61 Одна такая команда — OPERA, "открывшая" сверхсветовые нейтрино.

62 См., например, сайт Международного консорциума по психиатрической геномике: *https://pgc.unc.edu.*

63 Обсуждение см. здесь: Visscher P. M. et al. *10 Years of GWAS Discovery: Biology, Function, and Translation*. The American Journal of Human Genetics. 101, no. 1 (2017): 5–22. Конечно же, генетические исследования тоже подвержены мощному и неоправданному хайпу, см. Caulfield T. *Spinning the Genome: Why Science Hype Matters*. Perspectives in Biology and Medicine. 61, no. 4 (2018): 560–71.

64 Нейронаука: *http://enigma.ini.usc.edu;* эпидемиология рака: *https://epi.grants.cancer.gov/InterLymph/;* психология: *https://psysciacc.org;* трансляционные медицинские исследования: *www.ed.ac.uk/clinical-brain-sciences/research/camarades.* Кто-то из ученых уже занимается "состязательным сотрудничеством", воплощающим научную норму организованного скептицизма: они

пытаются противостоять своим предрассудкам, намеренно работая с исследователями, занимающими в каком-либо научном споре противоположную точку зрения. Идея в том, что если обе стороны согласятся подвергнуть теорию честной проверке и сообща эту проверку проведут, то результат для всех участников процесса будет убедительнее. Мой любимый пример состязательного сотрудничества — тот, где поборники существования экстрасенсорных способностей и скептики вместе проводили серию экспериментов, проверяя, способны ли люди "психически" определить, когда кто-то смотрит им в затылок. Не буду раскрывать результаты, поскольку статья стоит того, чтобы ее прочитать. Schlitz M. et al. *Of Two Minds: Sceptic-Proponent Collaboration within Parapsychology*. British Journal of Psychology. 97, no. 3 (2006): 313–22.

65 *www.usa.gov/government-works*

66 *www.coalition-s.org*

67 Else H. *Radical Open-Access Plan Could Spell End to Journal Subscriptions*. Nature. 561, no. 7721 (2018): 17–8. Должен отметить, что не все приветствуют открытый доступ. Пожалуй, лучший и самый подробный разбор его плюсов и минусов написал исследователь искусственного интеллекта Дэниел Аллингтон: Allington D. *On Open Access, and Why It's Not the Answer*. Daniel Allington. 15 Oct. 2013.

68 Schekman R. *Scientific Research Shouldn't Sit behind a Paywall*. Scientific American. 20 June 2019. Я решил упомянуть Калифорнийский университет и издательство *Elsevier* потому, что на момент написания этих строк они ведут ожесточенный спор: университет пытался заставить издательство уменьшить расценки, оно отказалось, и тогда университет принял суровые меры, попросту отменив свою подписку на журналы издательства *Elsevier*. Похвально. См. University of California Office of Scholarly Communication. *UC and Elsevier*. 20 March 2019.

69 Согласно одному отчету, по меньшей мере в какие-то годы научное издательство *Elsevier* по маржинальности чистой прибыли обошло *Apple*, *Google* и *Amazon*: Buranyi S. *Is the Staggeringly Profitable Business of Scientific Publishing Bad for Science?* The Guardian. 27 June 2017. Истинная нелепость позиции *Elsevier* в отношении научных публикаций проявилась в 2016 году, когда они запатентовали идею "онлайн-рецензирования" — и удостоились награды "Самый дурацкий патент месяца" за август 2016 года по версии организации *Electronic Frontier Foundation*. См. Nazer D., Harmon E. *Stupid Patent of the Month: Elsevier Patents Online Peer Review*. Electronic Freedom Foundation. 31 Aug. 2016. Еще о грехах *Elsevier* можно почитать тут: Yarkoni T. *Why I Still Won't Review for or Publish with Elsevier — and Think You Shouldn't Either*. [Citation Needed]. 12 Dec. 2016.

70 Buranyi S. *Is the Staggeringly Profitable Business of Scientific Publishing Bad for Science?*

71 Хотя бы издательство *Wiley*, в отличие от *Elsevier*, проявило интерес к переговорам о новых издательских моделях: Kwon D. *As Elsevier Falters, Wiley Succeeds in Open-Access Deal Making*. Scientist. 26 March 2019.

72 Препринты по физике можно найти на *arXiv* (*https://arxiv.org);* произносится "архив", поскольку X — это на самом деле греческая буква хи (χ). Препринты по экономике (там их обычно называют "рабочими документами") — в "Национальном бюро экономических исследований" (*www.nber.org/papers.html)* помимо других мест. Основной сервер препринтов по биологии — *bioRxiv* (*www.biorxiv.org);* по медицине — *medRxiv* (*www.medrxiv.org);* по психологии — *PsyArXiv* (*https://psyarxiv.com).* Вот список серверов для других дисциплин: *https://osf.io/preprints.* Описание того, насколько популярны стали препринты в биологии, а также насколько быстро, см. здесь: Abdill R. J., Blekhman J. *Tracking the Popularity and Outcomes of All bioRxiv Preprints*. eLife. 8 (2019): e45133. Еще вам может быть интересно, как серверы препринтов финансируются. Бывает, что за счет научных грантов, однако какие-то архивы финансируются по более устойчивой модели: например, у *arXiv*, самого первого, заключены соглашения с разными университетами, каждый из которых вносит свою финансовую лепту, чей размер зависит от того, насколько много ученые из этого университета скачивают и используют препринты с данного сервера (см. *https://arxiv.org/about/ourmembers).*

73 Как отмечено в примечании 55 к этой главе, некоторые стандартные журналы тоже делают открытым свой процесс рецензирования.

74 Эта идея в разных вариантах обсуждается в данных работах: Nosek B. A., Bar-Anan Y. *Scientific Utopia: I. Opening Scientific Communication*. Psychological Inquiry. 23, no. 3 (2012): 217–43; Stern B. M., O'Shea E. K. *A Proposal for the Future of Scientific Publishing in the Life Sciences*. PLOS Biology. 17, no. 2 (2019): e3000116. Также см. Birukou A. et al. *Alternatives to Peer Review: Novel Approaches for Research Evaluation*. Frontiers in Computational Neuroscience. 5 (2011).

75 Перспективное воплощение этой идеи — организация *Peer Community In* (PCI; *https://peercommunityin.org),* которая "стремится создать специальные сообщества исследователей, бесплатно рецензирующих и рекомендующих не опубликованные в журналах препринты в своей области". В PCI уже входят сообщества ученых в том числе из таких областей, как эволюционная биология, экология, палеонтология, нейронаука, геномика.

76 В данной работе даже предлагается, чтобы журналы курировали случаи "исследований, требующих повторения" и "статей, содержащих спорные утверждения": Stern B. M., O'Shea E. K. *A Proposal for the Future*.

77 А можно представить себе комбинацию этой модели с принципом "зарегистрированных отчетов", рассмотренным нами ранее, — когда предварительно описываются в препринтах и оцениваются не готовые работы, а планы исследований. Еще такая система могла бы использовать алгоритмы проверки ошибок, внедрить которые для репозитория препринтов не сложнее, чем для журнала старого образца.

78 Вообще говоря, в небольшом масштабе кое-что уже тестируется: *https://asapbio.org/eisen-appraise.*

79 Следует отметить, что в большинстве архивов с препринтами работы просматриваются, чтобы не допустить появления дичайшей ерунды, размещаемой злоумышленниками или безумцами. Однако полноценное рецензирование не осуществляется (да это и невозможно, учитывая, сколько на все про все уходило бы времени).

80 Fryer R. *An Empirical Analysis of Racial Differences in Police Use of Force.* Working Paper no. 22399. National Bureau of Economic Research. 2016.

81 Там же.

82 Bui Q., Cox A. *Surprising New Evidence Shows Bias in Police Use of Force but Not in Shootings*. New York Times. 11 July 2016.

83 Elder L. *Ignorance of Facts Fuels the Anti-Cop "Movement"*. RealClear Politics. 14 July 2016.

84 Simonsohn U. *Teenagers in Bikinis: Interpreting Police-Shooting Data*. Data Colada. 14 July 2016.

85 Этот факт был отмечен в строке препринта, следующей за сообщением о значении 23,8 %, но в ходе общения Фрайера с прессой, по-видимому, не всплывал. Дополнительную критику исследования Фрайера можно найти в следующих блогах экономистов: Rajiv S. *Police Use of Force: Notes on a Study.* Blog. 11 July 2016; Feldman J. *Roland Fryer is Wrong: There is Racial Bias in Shootings by Police*. Harvard University, blog. 12 July 2016.

86 Fryer R. *An Empirical Analysis of Racial Differences in Police Use of Force*. Journal of Political Economy. 127, no. 3 (2019): 1210–61.

87 Кстати, препринт Фрайера несколько его коллег-экономистов просмотрели, просто он не был отрецензирован по форме и опубликован. Engber D. *Was This Study Even Peer-Reviewed?* Slate. 25 July 2016.

88 *www.biorxiv.org/content/early/recent.* (Обратите внимание: предупреждение скорее временное, так что к моменту выхода этой книги его уже могут изменить или удалить.)

89 Kupferschmidt K. *Preprints Bring "Firehose" of Outbreak Data*. Science. 367, no. 6481 (2020): 963–4.

90 McKiernan E. et al. *The "Impact" of the Journal Impact Factor in the Review, Tenure, and Promotion Process*. Impact of Social Sciences. 26 April 2019.

91 Flier J. S. *Credit and Priority in Scientific Discovery: A Scientist's Perspective*. Perspectives in Biology and Medicine. 62, no. 2 (2019): 189–215.

92 Совсем иначе обстоит дело в некоторых других областях, таких как математика и в меньшей степени физика, где авторы указываются по алфавиту (хотя такая практика, судя по всему, становится все менее распространенной). См. Waltman L. *An Empirical Analysis of the Use of Alphabetical Authorship in Scientific Publishing*. Journal of Informetrics. 6, no. 4 (2012): 700–11.

93 Mallapaty S. *Paper Authorship Goes Hyper*. Nature Index. 30 Jan. 2018.

94 Longo D. L., Drazen J. M. *Data Sharing*. New England Journal of Medicine. 374, no. 3 (2016): 276–7. Еще редакторы журнала *The New England Journal of Medicine* беспокоились, как бы исследователи-паразиты не стали "даже использовать данные для того, чтобы пытаться опровергнуть установленное первоначальными исследователями", что уж явно нелепо. Однако есть сходное опасение, которое обсудить стоит: если ваши данные открыты и вы публикуете в интернете свои планы относительно того, что собираетесь с ними делать, вас могут обойти — другие ученые возьмут ваши данные, проведут ваш анализ и поспешат опубликовать статью раньше вас. В отрасли, столь ориентированной на новизну, карьеру вот так обойденного ученого может ждать крах. Впрочем, неясно, насколько часто это происходит, и есть несколько соображений, которые нужно изложить. Во-первых, в деликатных ситуациях вы вправе наложить запрет на предварительные регистрации — чтобы они стали достоянием общественности только после того, как вы проведете анализ данных (но при этом есть отметка времени, показывающая, когда вы зарегистрировали свои гипотезы). Во-вторых, обидно, конечно, оказаться обойденным, но на самом деле не так уж плохо, если разные люди проводят один и тот же анализ одних и тех же данных: сравнивая разные версии, можно выявить ошибки или упущения. Именно поэтому как минимум один журнал *(PLOS Biology)* позволяет обойденным исследователям подавать на рассмотрение собственную статью, у которой все равно будут все шансы на публикацию: The PLOS Biology Staff Editors. *The Importance of Being Second*. PLOS Biology. 16, no. 1 (2018): e2005203.

95 Одна группа исследователей предложила использовать *T*-индекс — показатель, пересчитывающий индекс Хирша автора на основании того, какой реальный вклад тот внес в каждую работу, на которой стоит его имя. Если ученый склонен возглавлять большинство своих проектов, то его *T*-индекс будет выше; если же он чаще лишь поддерживает работу других, то показатель в итоге будет ниже.

Однако для этого все ученые должны прежде всего честно признаваться, каков объем выполненной ими работы для каждой статьи, что, вероятно, утопия. И конечно, такой показатель все равно можно подтасовать, поэтому никогда нельзя будет ориентироваться исключительно на него, как и в случае с индексом Хирша. Rahman M. T. et al. *The Need to Quantify Authors' Relative Intellectual Contributions in a Multi-Author Paper*. Journal of Informetrics. 11, no. 1. (2017): 275–81. См. также Shen H. W., Barabási A. L. *Collective Credit Allocation in Science*. Proceedings of the National Academy of Sciences. 111, no. 34 (2014): 12325–30.

96 Один пример см. тут: Moher D. et al. *Assessing Scientists for Hiring, Promotion, and Tenure*. PLOS Biology. 16, no. 3 (2018): e2004089. Сан-Францисскую декларацию об оценке научных исследований, в которой приводятся убедительные аргументы против использования допускающих подтасовку показателей при найме ученых и оценке исследований, можно найти здесь: *https://sfdora.org.*

97 Naudet F. et al. *Six Principles for Assessing Scientists for Hiring, Promotion, and Tenure*. Impact of Social Sciences. 4 June 2018.

98 Lilienfeld S. O. *Psychology's Replication Crisis and the Grant Culture: Righting the Ship*. Perspectives on Psychological Science. 12, no. 4 (2017): 661–4.

99 Ioannidis J. P. A. *Fund People Not Projects*. Nature. 477, no. 7366 (2011): 529–31; Wilkinson E. *Wellcome Trust to Fund People Not Projects*. Lancet. 375, no. 9710 (2010): 185–6.

100 Fang F. C., Casadevall A. *Research Funding: The Case for a Modified Lottery*. MBio. 7, no. 2 (2016): 5.

101 Fang F. C. et al. *NIH Peer Review Percentile Scores Are Poorly Predictive of Grant Productivity*. eLife. 5 (2016): e13323.

102 Gross K., Bergstrom C. T. *Contest Models Highlight Inherent Inefficiencies of Scientific Funding Competitions*. PLOS Biology. 17, no. 1 (2019): e3000065.

103 Bishop D. *Luck of the Draw*. Nature Index. 7 May 2018. Также см. Vazire S. *Our Obsession with Eminence Warps Research*. Nature. 547, no. 7661 (2017): 7.

104 Smaldino P. et al. *Open Science and Modified Funding Lotteries Can Impede the Natural Selection of Bad Science*. Open Science Framework. Preprint (29 Jan. 2019). *https://doi.org/10.31219/osf.io/zvkwq*

105 *American Journal of Political Science* не только требует, чтобы исследователи делились своими данными, но и в процессе рецензирования каждой статьи явным образом проверяет, что результаты воспроизводимы. См.: *https://ajps.org/ajps-verification-policy.*

106 Nosek B. A. et al. *Promoting an Open Research Culture*.

107 Существуют также своды правил, под которыми подписались многие журналы. Таковы, например, инструкции Комитета по публи-

кационной этике (COPE, *Committee on Publication Ethics*; *https://publicationethics.org/guidance/Guidelines)* — предоставляемый редакторам набор лучших рекомендаций по борьбе с недобросовестной научной практикой. Зачастую проблема лишь в том, как добиться, чтобы редакторы действительно следовали этим рекомендациям.

108 Есть свидетельства того, что включение в пресс-релизы предостерегающих заявлений улучшает освещение научных результатов в средствах массовой информации. В обсуждавшемся в шестой главе рандомизированном контролируемом исследовании, когда ученые меняли содержание пресс-релизов, обнаружилось следующее: при добавлении к описанию результатов предупреждения о том, что “утверждения корреляционные, а не причинно-следственные”, 20 % новостных материалов это подхватывали, тогда как ранее подобного заявления почти ни в одном из них не было. И пусть ученые не беспокоятся, что их пресс-релизами, если включить туда подобные оговорки, заинтересуется меньше журналистов: в том же исследовании не обнаружилось никаких доказательств того, что осторожно написанные пресс-релизы (где меньше преувеличений) реже преобразуются в новостные статьи, хотя ранее подобное отмечалось в нескольких других работах, посвященных пресс-релизам. Adams R. C. et al. *Claims of Causality in Health News*. См. также Sumner P. et al. *Exaggerations and Caveats in Press Releases and Health-Related Science News*. PLOS ONE. 11, no. 12 (2016): e0168217; Bott L. et al. *Caveats in Science-Based News Stories Communicate Caution without Lowering Interest*. Journal of Experimental Psychology: Applied. 25, no. 4 (2019): 517–42. В случае со “сверхсветовыми” нейтрино мы видели, как осторожно сформулированный пресс-релиз преобразовался в двусмысленные новостные заметки.

109 Tiokhin L. et al. *Honest Signaling in Academic Publishing.* Open Science Framework. Preprint (13 June 2019). *https://doi.org/10.31219/osf.io/gyeh8*

110 См., например, вот эту редакционную статью из журнала *Psychological Science*: Eich E. *Business Not as Usual*. Psychological Science. 25, no. 1 (2014): 3–6.

111 Munafo M. *Raising Research Quality Will Require Collective Action*. Nature. 576, no. 7786 (2019): 183. Есть много аналогичных инициатив, таких как “Сообщества открытой науки” в Нидерландах (*https://osf.io/vz2sy)* и Центр открытой науки в Мюнхенском университете имени Людвига и Максимилиана (*www.osc.uni-muenchen.de/index.html)*. Еще одно успешное общественное движение — *ReproducibiliTea*, которое устраивает встречи “журнального клуба” для начинающих исследователей, где обсуждаются вопросы, связанные с открытой наукой (*https://reproducibilitea.org)*.

112 Hardwicke T. E. et al. *Calibrating the Scientific Ecosystem Through Meta-Research*. Annual Review of Statistics and Its Application. 7, no. 1 (2020): 11–37.

113 Конечно, это может зайти и слишком далеко. Случай Эми Кадди, о котором мы уже говорили, — это один из примеров, как обсуждение в интернете, независимо от того, насколько плохи были научные данные, приобретает непропорциональный, издевательский и почти ликующий характер. См. Dominus S. *When the Revolution Came for Amy Cuddy*. New York Times. 18 Oct. 2017. Кстати, упоминание об издевательствах — буллинге — поднимает вопрос, выходящий за рамки данной книги, но все же заслуживающий внимания. Одним из основных аспектов науки, который тоже остро нуждается в реформировании, является культура власти. Существует слишком много (хотя даже одна — это уже слишком много) историй о том, как научные сотрудники злоупотребляют своим положением старших по отношению к студентам или другим исследователям, чтобы издеваться над ними, приставать или даже осуществлять сексуальное насилие. Два недавних случая из области психологии иллюстрируют это удручающее явление. Специалист по социальной нейронауке Таня Сингер в 2018 году ушла с поста директора Института когнитивных наук и наук о человеческом мозге Общества Макса Планка в Лейпциге после обвинений в том, что она годами жестоко издевалась над своей научной группой; например, рассказывали, как она орала на забеременевшую молодую сотрудницу (потому что ее декретный отпуск прервал бы исследования Сингер). Ирония ситуации заключалась в том, что главный научный интерес Сингер — человеческая эмпатия (Kupferschmidt K. *She's the World's Top Empathy Researcher. But Colleagues Say She Bullied and Intimidated Them*. Science. 8 Aug. 2018). А в Дартмутском колледже три профессора психологии — Тодд Хизертон, Уильям Келли и Пол Вэйлен — лишились работы и получили запрет на посещение кампуса (или ограничение доступа туда) после того, как девять женщин выступили с обвинениями против них в домогательствах, сексуальном насилии и даже изнасилованиях студенток на протяжении шестнадцати лет. The Dartmouth Senior Staff. *New Allegations of Sexual Assault Made in Ongoing Lawsuit against Dartmouth*. The Dartmouth. 2 May 2019.

114 Nosek B. *Strategy for Culture Change*. Center for Open Science. 11 July 2019.

115 Markowetz F. *Five Selfish Reasons to Work Reproducibly*. Genome Biology. 16:274 (2015).

116 Robin W. *How a Somber Symphony Sold More Than a Million Records*. New York Times. 9 June 2017. Особенно хороша запись симфонического оркестра Польского национального радио, сделанная

в 2019 году (дирижировал Кшиштоф Пендерецкий) и выпущенная компанией *Domino Recording Co Ltd*.: *https://open.spotify.com/album/6r4bpBHOQzQ8oJoYmzmKZK*.

117 Howard L. B. *Henry M. Górecki's Symphony No. 3 (1976) As A Symbol of Polish Political History*. Polish Review. 52, no. 2 (2007): 215–22.

118 Это различие заставляет вспомнить о старом споре в эволюционной биологии: происходит ли эволюция неуклонно и постепенно ("градуализм"), или же она в основном статична, но прерывается внезапным всплескообразным появлением новых видов ("прерывистое равновесие"). См. Sterelny K. *Dawkins vs. Gould: Survival of the Fittest*. Thriplow: Icon Books, 2007.

119 Когда мы разрешим все проблемы, описанные в этой книге, настанет время заняться более серьезными вопросами. В конечном итоге мы хотим встроить наши научные результаты в сильные теории, которые объясняют мир и предсказывают будущие наблюдения. См. Muthukrishna M., Henrich J. *A Problem in Theory*. Nature Human Behaviour. 3, no. 3 (2019): 221–9. Но в мире, где любой результат при попытке его воспроизвести рискует развалиться на части, сложные теории могут в итоге завести нас совершенно не туда. См. Deary I. J. *Looking Down on Human Intelligence: From Psychometrics to the Brain* // Oxford Psychology Series. No. 34. Oxford: Oxford University Press, 2000 (особенно с. 108–109). Более приземленная цель, чем построение теорий, но все же заслуживающая внимания — это *триангуляция*, когда к вопросу подходят с нескольких разных сторон, используя разного рода исследования с различными исходными допущениями и проверяя, сходятся ли они все к единому ответу. См. Munafo M. R., Smith G. D. *Robust Research Needs Many Lines of Evidence*. Nature. 553, no. 7689 (2018): 399–401; Lawlor D. A. et al. *Triangulation in Aetiological Epidemiology*. International Journal of Epidemiology. 45, no. 6 (2017): 1866–86. Исторический пример триангуляции: Smith G. D. *Smoking and Lung Cancer: Causality, Cornfield and an Early Observational Meta-Analysis*. International Journal of Epidemiology. 38, no. 5 (2009): 1169–71. Но опять же: если мы не можем полагаться на отдельные результаты, то пытаться использовать их для триангуляции — дохлый номер.

120 Классический пример — открытие зеленого флуоресцентного белка (GFP), который ярко флуоресцирует под ультрафиолетовым светом и в настоящее время широко используется в биологии как метка, сообщающая о присутствии в клетке определенных белков. Это было чрезвычайно важным достижением, открывшим множество новых исследовательских направлений в биологии, однако начиналось все очень скромно: Осаму Шимомура (который за открытие зеленого флуоресцентного белка в 2008 году совместно

с другими учеными получил Нобелевскую премию по химии) обнаружил GFP в 1960-х годах, выделяя белки из особой биолюминесцентной медузы в рамках проекта, изначально казавшегося тупиковым. См. *Osamu Shimomura. Biographical*. NobelPrize.org. 2008. Другие примеры того, как прожектерские исследования в конце концов неожиданно приводят к крупным достижениям, обсуждаются здесь: Bhattacharya J., Packalen M. *Stagnation and Scientific Incentives*. Working Paper no. 26752. National Bureau of Economic Research. 2020.

121 Bhattacharya J., Packalen M. *Stagnation and Scientific Incentives.*

122 Nuijten M. B. et al. *Practical Tools and Strategies for Researchers to Increase Replicability*. Developmental Medicine & Child Neurology. 61, no. 5 (2018): 535–9.

123 См., например, Fanelli D. *Opinion: Is Science Really Facing a Reproducibility Crisis, and Do We Need It To?* Proceedings of the National Academy of Sciences. 115, no. 11 (2018): 2628–31.

Эпилог

1 *https://eventhorizontelescope.org*

2 Тяжелый иммунодефицит: Mamcarz E. et al. *Lentiviral Gene Therapy Combined with Low-Dose Busulfan in Infants with SCID–X1.* New England Journal of Medicine. 380, no. 16 (2019): 1525–34. Муковисцидоз: Collins F. S. *Realizing the Dream of Molecularly Targeted Therapies for Cystic Fibrosis*. New England Journal of Medicine. 381, no. 19 (2019): 1863–5.

3 Rodger A. J. et al. *Risk of HIV Transmission through Condomless Sex in Serodifferent Gay Couples with the HIV-Positive Partner Taking Suppressive Antiretroviral Therapy (PARTNER): Final Results of a Multicentre, Prospective, Observational Study*. Lancet. 393, no. 10189 (2019): 2428–38.

4 Tsurumoto K. et al. *Quantum Teleportation-Based State Transfer of Photon Polarization into a Carbon Spin in Diamond*. Communications Physics. 2, no. 1 (2019): 74.

5 Ma Y. et al. *Mammalian Near-Infrared Image Vision through Injectable and Self-Powered Retinal Nanoantennae*. Cell. 177, no. 2 (2019): 243–55.

6 Handley E. A. *Findings of Research Misconduct*. Federal Register. 84, no. 216 (2019): 60097–8. Предысторию см. здесь: McCook A. $ *200M Research Misconduct Case against Duke Moving Forward, as Judge Denies Motion to Dismiss*. Retraction Watch. 28 April 2017.

7 Sample I. *Top Geneticist "Should Resign" Over His Team's Laboratory Fraud*. The Guardian. 1 Feb. 2020.

8 Исходная статья: Oxley D. R. et al. *Political Attitudes Vary with Physiological Traits*. Science. 321, no. 5896 (2008): 1667–70. Авторы ис-

следования-повторения описали всю историю: Arceneaux K. et al. *We Tried to Publish a Replication of a Science Paper in Science. The Journal Refused*. Slate. 20 June 2019. Их статья о повторении той работы в итоге была опубликована как Bakker B. N. et al. *Conservatives and Liberals Have Similar Physiological Responses to Threats*. Nature Human Behaviour. 4 (2020): 613–21.

9 Chawla D. S. *Russian Journals Retract More than 800 Papers after "Bombshell" Investigation*. Science. 8 Jan. 2020.

10 Noorden R. V. *Highly Cited Researcher Banned from Journal Board for Citation Abuse*. Nature. 578, no. 7794 (2020): 200–1.

11 Rennie D. *Guarding the Guardians: A Conference on Editorial Peer Review*. JAMA. 256, no. 17 (1986): 2391.

12 Babbage C. *Reflections on the Decline of Science in England, and on Some of Its Causes*. London: B. Fellowes, 1830.

13 *International Biographical Dictionary of Computer Pioneers*. Chicago: Fitzroy Dearborn, 1995.

14 Chaplin S. et al. *Wellcome Trust Global Monitor 2018*. Wellcome Trust. 19 June 2019.

15 Опрос компании *Ipsos MORI* о доверии к науке показал, что в Великобритании доля людей, верящих, что ученые говорят правду, в 2018 году составила 85 %, что на 22 % больше, чем в 1997-м, когда этот опрос начали проводить. Skinner G., Clemence M. *Ipsos MORI Veracity Index 2018*. Ipsos MORI. Nov. 2018.

16 Пример из психологии см. здесь: Anvari F., Lakens D. *The Replicability Crisis and Public Trust in Psychological Science*. Comprehensive Results in Social Psychology. 3, no. 3 (2018): 266–86. В другом исследовании, посвященном тому, как люди реагируют, когда им открыто заявляют, что в науке случались провалы при попытке воспроизвести исследования, только 17 % из (признаться, малой) выборки согласились, что это повод не доверять науке в целом. См. Weißkopf M. et al. *Wissenschaftsbarometer 2018*. Wissenschaft im Dialog. 2018.

17 Насколько я знаю, исходно это выражение Джона Даймонда. Цитируется, например, тут: Jeffery N. *"There Is No Such Thing as Alternative Medicine"*. Journal of Small Animal Practice. 56, no. 12 (2015): 687–8.

18 Csiszar A. *The Scientific Journal: Authorship and the Politics of Knowledge in the Nineteenth Century*. Chicago: University of Chicago Press, 2018.

19 Guarino B. *USDA Orders Scientists to Say Published Research Is "Preliminary"*. Washington Post. 19 April 2019. Министерство сельского хозяйства США также обвинили в политически мотивированной, скажем так, шумихе наоборот: оно отменило стандартное правило публиковать результаты исследований, проведенных учеными министерства,

в тех случаях, когда эти исследования подчеркивали опасность изменения климата. Evich H. B. *Agriculture Department Buries Studies Showing Dangers of Climate Change*. Politico. 23 June 2019.

20 Guarino B. *After Outcry, USDA Will No Longer Require Scientists to Label Research "Preliminary"*. Washington Post. 10 May 2019.

21 Marcus A., Oransky I. *Trump Gets Something Right about Science, Even If for the Wrong Reasons*. Washington Post. 1 May 2019.

22 О Лысенко неплохо написано тут: Grant J. *Corrupted Science: Fraud, Ideology and Politics in Science*. London: Facts, Figures & Fun, 2007. О Сталине и Мао: Kean S. *The Soviet Era's Deadliest Scientist Is Regaining Popularity in Russia*. The Atlantic. Dec. 2017. Описание недавнего возрождения идей Лысенко в России, вызывающее тревогу, см. здесь: Kolchinsky E. I. et al. *Russia's New Lysenkoism*. Current Biology. 27, no. 19 (2017): R1042–7.

23 Креационизм: Devi G. *Creationism Isn't Just an Ideology — It's a Weapon of Political Control*. The Guardian. 22 Nov. 2015. Вакцины (пожалуй, иронично, что в 2019 году один из главных итальянских политиков-антипрививочников был госпитализирован с ветряной оспой): Kington T. *Italian "Anti-Vax" Advocate Massimiliano Fedriga Catches Chickenpox*. The Times. 20 March 2019. ВИЧ и СПИД: Chigwedere P. et al. *Estimating the Lost Benefits of Antiretroviral Drug Use in South Africa*. JAIDS. 49, no. 4 (2008): 410–5. Технология использования стволовых клеток: Sohini C. *Bowel Cleanse for Better DNA: The Nonsense Science of Modi's India*. South China Morning Post. 13 Jan. 2019.

24 "Чистый и зеленый": Scottish National Party. *Why Have the Scottish Government Banned GM Crops?* "Дешевый популизм": McColm E. *Ban on GM crops is embarrassing*. The Scotsman. 18 Aug. 2015. "Чрезвычайно тревожная": Stokstad E. *Scientists Protest Scotland's Ban of GM Crops*. Science. 17 Aug. 2015.

25 Zola É. *Proudhon et Courbet*. 1866. Цитируется и переведено в Dorra H. *Symbolist Art Theories: A Critical Anthology*. Berkeley: University of California Press, 1994. [Золя Э. *Прудон и Курбе* // Собр. соч. в 26-ти т. Т. 24. М.: Художественная литература, 1966.]

Приложение. **Как читать научную статью**

1 Himmelstein D. S. *Sci-Hub Provides Access to Nearly All Scholarly Literature*. eLife. 7 (2018): e32822.

2 Еще можно свериться с одним из опубликованных в интернете списков хищнических журналов, например, здесь: *https://beallslist.net.*

3 Вот тут для многих стран и регионов можно найти ссылки на сайты, где регистрируются клинические испытания: *www.hhs.gov/ohrp/international/clinical-trial-registries/index.html.* Для других видов иссле-

дований рекомендую проверять сайты вроде таких: *https://arxiv.org; www.biorxiv.org; https://osf.io.* Многие предварительно зарегистрированные статьи будут содержать ссылку на веб-страницу с их регистрацией, а клинические испытания будут иметь свой регистрационный номер, который вы сможете использовать.

4 Часто ссылку на полный набор данных можно найти в конце опубликованной работы. Еще некоторые журналы сейчас обозначают статьи с открытыми данными и методами и предварительной регистрацией цветными плашками. Список таких журналов можно посмотреть здесь: *https://cos.io/our-services/open-science-badges.*

5 Это идея моей подруги Салони Даттани.

6 Функция “Цитируется” в *Google Scholar* очень удобно расположена под записью о каждой статье.

7 *www.sciencemediacentre.org.* “Научный пресс-центр” есть и в некоторых других странах, например в Германии: *www.sciencemediacenter.de.* Подробнее см. Callaway E. *Science Media: Centre of Attention*. Nature. 499, no. 7457 (2013): 142–4.

8 *https://pubpeer.com*

9 Можно просто вставить URL-адрес журнальной статьи в поисковую строку твиттера — так вы увидите все сообщения со ссылками на эту статью и все комментарии к ним. Многие ученые стараются делать критические обзоры статей, относящихся к их области исследования, в доступной форме в твиттере, так что это источник комментариев для широкой публики, который мало кто использует.

10 Интересный новый инструмент, позволяющий лучше понять, как цитируется то или иное исследование, — *scite* (*https://scite.ai*). Алгоритм (при его обучении использовались суждения реальных ученых) анализирует контекст, в котором упоминается цитируемая статья, и классифицирует его как “подтверждающий”, “противоречащий” или просто “упоминающий” (то есть нейтральный). Хотя алгоритм еще дорабатывается и функционирует небезупречно, он предоставляет отрывки текста из каждой статьи, где цитируется интересующая вас работа, чтобы вы могли сделать собственные выводы. Это лишь один пример технологических инструментов, которые постоянно появляются, чтобы облегчить жизнь ученым — и предотвратить их ошибки.

Послесловие

1 *Moderna Provides Business Update and Announces Three New Development Programs in Infectious Disease Vaccines.* Moderna. 11 Jan. 2021.

2 В начале 2021 года мы с несколькими моими друзьями создали сайт “Антивирус”, где собрали часто задаваемые вопросы о COVID-19,

чтобы разоблачить самую вредоносную ложную информацию, распространявшуюся во время пандемии.

3 Я писал об истории с гидроксихлорохином, а также о массе других поучительных историй эпохи коронавирусной пандемии здесь: Ritchie S. *The Great Reinforcer.* Works in Progress. 3 (2021).

4 Gautret P. et al. *Hydroxychloroquine and Azithromycin as a Treatment of* COVID-*19: Results of an Open-label Non-randomized Clinical Trial*. International Journal of Antimicrobial Agents. 56, no. 5 (2020): 105949.

5 Макрон: *Macron Visits Marseille Doctor Behind Hydroxychloroquine Coronavirus "Cure" Touted by Trump.* France 24. 9 Apr. 2020. Трамп: *Trump Announces Potential "Game Changer" on Drugs to Treat Novel Coronavirus, But FDA Says More Study is Needed.* ABC News. 19 Mar. 2020.

6 Wong J. C. *Hydroxychloroquine: How an Unproven Drug Became Trump's Coronavirus "Miracle Cure".* The Guardian. 7 Apr. 2020.

7 Еще букет проблем со статьей обнаружила Элизабет Бик (мы говорили о ней в третьей главе): Bik E. *Thoughts on the Gautret et al. Paper About Hydroxychloroquine and Azithromycin Treatment of* COVID-*19 Infections.* Science Integrity Digest. 24 Mar. 2020. Также см. Hulme O. J. et al. *A Bayesian Reanalysis of the Effects of Hydroxychloroquine and Azithromycin on Viral Carriage in Patients with* COVID-*19*. PLOS ONE. 16, no. 2 (2021): e0245048.

8 Glasziou P. P. et al. *Waste in Covid-19 Research*. BMJ. 369 (2020): m1847.

9 Пример — масштабное высококачественное исследование *RECOVERY Trial*. The RECOVERY Collaborative Group. *Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19.* New England Journal of Medicine. 383 (2020): 2030–40.

10 Mehra M. R. et al. *Hydroxychloroquine or Chloroquine with or without a Macrolide for Treatment of* COVID-*19: A Multinational Registry Analysis.* Lancet. 22 May 2020.

11 Всемирная организация здравоохранения: Beaubien J. *WHO Halts Hydroxychloroquine Over Safety Concerns.* NPR Coronavirus Updates. 25 May 2020. Франция: Ledsom A. *France Says No to Hydroxychloroquine Prescription After Lancet Study.* Forbes. 26 May 2020.

12 Davey M. et al. *Surgisphere: Governments and WHO Changed Covid-19 Policy Based on Suspect Data from Tiny US Company*. The Guardian. 3 Jun. 2020.

13 Все проблемы подробнейше изложены на сайте с дискуссиями по науке *PubPeer: https://pubpeer.com/publications/71DA593B9943638F1ADE6F80696914.*

14 Mehra M. R. et al. *Retraction — Hydroxychloroquine or Chloroquine with or Without a Macrolide for Treatment of* COVID-*19: A Multinational Registry Analysis*. Lancet. 395, no. 10240 (2020): 1820. В то же

время и *The New England Journal of Medicine* отозвал статью, основанную на данных *Surgisphere*: Mehra M. R. et al. *Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19.* New England Journal of Medicine. 382 (2020): e102.

15 The Editors of the Lancet Group. *Learning From a Retraction.* Lancet. 396, no. 10257 (2020): 1056.

16 *French Professor Faces Disciplinary Case Over Hydroxychloroquine Claims.* The Guardian. 12 Nov. 2020.

17 O'Grady C. *Journals Singled Out for Favoritism.* Science. 25 Feb. 2021. Профиль Рауля на *Google Scholar* с его индексом Хирша и другими сведениями: *https://scholar.google.co.uk/citations?user=n8EF_6kAAAAJ.*

18 О многих других случаях проведения низкокачественных исследований во время пандемии, а также о том, как можно было их избежать, см. здесь: Besançon L. et al. *Open Science Saves Lives: Lessons from the* COVID-*19 Pandemic*. BioRxiv (30 Oct. 2020).

19 Marcus A. *Researcher to Overtake Diederik Stapel on the Retraction Watch Leaderboard, with 61.* Retraction Watch. 2 Feb. 2021.

20 Frolov S. *Quantum Computing's Reproducibility Crisis: Majorana Fermions.* Nature. 592 (2021): 350–2.

21 Friese M., Frankenbach J. *p-Hacking and Publication Bias Interact to Distort Meta-analytic Effect Size Estimates*. Psychological Methods. 25, no. 4 (2020): 456–71.

22 Marcus A. *Elsevier Looking into "Very Serious Concerns" after Student Calls out Journal for Fleet of Star Trek Articles, Other Issues*. Retraction Watch. 10 Dec. 2020.

CORPUS 798

Научно-популярное издание

СТЮАРТ РИЧИ

НАУКО-ОБРАЗНАЯ ЧУШЬ

РАЗОБЛАЧЕНИЕ МОШЕННИЧЕСТВА, ПРЕДВЗЯТОСТИ, НЕДОБРОСОВЕСТНОСТИ И ХАЙПА В НАУКЕ

Главный редактор Варвара Горностаева

Художник Андрей Бондаренко

Редактор Екатерина Владимирская

Ответственный за выпуск Ольга Энрайт

Технический редактор Татьяна Полонская

Корректор Лилия Цинман

Верстка Марат Зинуллин

Общероссийский классификатор продукции
ОК-034-2014 (КПЕС 2008);
58.11.1 — книги, брошюры печатные

Подписано в печать 15.10.2023. Формат 60 × 90 1/16
Бумага офсетная. Гарнитура *OriginalGaramondC*
Печать офсетная. Усл. печ. л. 27,0
Тираж 2000 экз. Заказ №

Отпечатано в соответствии с предоставленными материалами
в ООО "ИПК Парето-Принт", 170546, Тверская область,
Промышленная зона Боровлево-1, комплекс №3А
www.pareto-print.ru

Произведено в Российской Федерации в 2024 г.
Изготовитель — ООО “Издательство АСТ”

ООО “Издательство АСТ”
129085, г. Москва, Звездный бульвар, дом 21, строение 1,
комната 705, пом. I, 7 этаж
Контактный адрес электронной почты: ask@ast.ru

“Баспа Аста” деген ООО
129085, Мәскеу қ., Звёздный бульвары, 21-үй, 1-құрылыс,
705-бөлме, I жай, 7-қабат
Біздің электрондық мекенжайымыз: ask@ast.ru

Интернет-магазин: www.book24.kz
Импортер в Республику Казахстан ТОО “РДЦ-Алматы”
Дистрибьютор и представитель по приему претензий
на продукцию в Республике Казахстан: ТОО “РДЦ-Алматы”

Интернет-дүкен: www.book24.kz
Қазақстан Республикасындағы импорттаушы “РДЦ-Алматы” ЖШС
Қазақстан Республикасында дистрибьютор және өнім бойынша
арыз-талаптарды қабылдаушының өкілі “РДЦ-Алматы” ЖШС
050039 Алматы қ., Домбровский көш., 3 “а”, литер Б, офис 1
Тел.: +7 (727) 251-59-89, 90, 91, 92, факс: +7 (727) 251-58-12, доб. 107
E-mail: RDC-Almaty@eksmo.kz
Өнімнің жарамдылық мерзімі шектелмеген

По вопросам оптовой покупки книг обращаться по адресу:
123112 г. Москва, Пресненская наб., д. 6, строение 2, БЦ “Империя”
Тел.: +7 (499) 951-60-00, доб. 574
E-mail: opt@ast.ru

